**UNIVERSIDADE DE SÃO PAULO
FACULDADE DE FILOSOFIA, LETRAS E CIÊNCIAS HUMANAS
DEPARTAMENTO DE LETRAS MODERNAS**

ELISA DUARTE TEIXEIRA

**A Lingüística de Corpus a serviço do tradutor:
proposta de um dicionário de Culinária
voltado para a produção textual**

São Paulo
2008

ELISA DUARTE TEIXEIRA

**A Lingüística de Corpus a serviço do tradutor:
proposta de um dicionário de Culinária
voltado para a produção textual**

Tese apresentada ao Programa Pós-Graduação em Estudos Lingüísticos e Literários em Inglês do Departamento de Letras Modernas da Faculdade de Filosofia, Letras e Ciências Humanas da Universidade de São Paulo para obtenção do título de Doutora em Letras.

Orientadora: Profa. Dra. Stella Esther Ortweiler Tagnin
Co-orientador: Prof. Dr. Antônio Paulo Berber Sardinha

São Paulo
2008

AUTORIZO A REPRODUÇÃO E DIVULGAÇÃO TOTAL OU PARCIAL DESTE TRABALHO, POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO, PARA FINS DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.

Catálogo na Publicação
Serviço de Biblioteca e Documentação
Faculdade de Filosofia, Letras e Ciências Humanas da Universidade de São Paulo

TEIXEIRA, Elisa Duarte

A Lingüística de Corpus a serviço do tradutor: proposta de um dicionário de culinária voltado para a produção textual / Elisa Duarte Teixeira; orientadora Stella Esther Ortweiler Tagnin; co-orientador Antônio Paulo Berber Sardinha -- São Paulo, 2008.

400 f. : fig., tab. + anexos

Tese (Doutorado – Programa de Pós-Graduação em Estudos Lingüísticos e Literários em Inglês do Departamento de Letras Modernas) – Faculdade de Filosofia, Letras e Ciências Humanas da Universidade de São Paulo.

1. Tradução. 2. Dicionários. 3. Culinária. 4. Lingüística aplicada. 5. Terminologia. 6. Lingüística computacional. I. Título. II. Tagnin, Stella Esther Ortweiler. III. Berber Sardinha, Antônio Paulo.

FOLHA DE APROVAÇÃO

Elisa Duarte Teixeira

A Lingüística de Corpus a serviço do tradutor:

proposta de um dicionário de Culinária voltado para a produção textual

Tese apresentada ao Programa Pós-Graduação em Estudos Lingüísticos e Literários em Inglês do Departamento de Letras Modernas da Faculdade de Filosofia, Letras e Ciências Humanas da Universidade de São Paulo para obtenção do título de Doutora em Letras.

Aprovada em: 01 / 12 / 2008

Banca examinadora

Profa. Dra. Stella Esther Ortweiler Tagnin

Instituição: USP

Assinatura: _____

Profa. Dra. Maria José Bocorny Finatto

Instituição: UFRGS

Assinatura: _____

Prof. Dr. Adauri Brezolin

Instituição: UESP

Assinatura: _____

Profa. Dra. Ieda Maria Alves

Instituição: USP

Assinatura: _____

Profa. Dra. Paola Giustina Baccin

Instituição: USP

Assinatura: _____

*Dedico este trabalho à Carol,
à minha família, e a todos os meus amigos do peito –
seu carinho é o que dá sentido às demais coisas da minha vida.*

AGRADECIMENTOS

Antes de mais, como diriam nossos irmãos portugueses, gostaria de agradecer à minha orientadora, Profa. Dra. Stella Tagnin, e ao meu co-orientador, Prof. Dr. Tony Berber Sardinha, pela amizade, pela confiança e, especialmente, pela paciência que tiveram comigo ao longo deste prolongado percurso – meu respeito ao trabalho de vocês foi o que me ajudou a continuar na senda quando a minha vontade era apenas de trilhar outros novos rumos. Agradeço também aos Profs. Drs. Paola Giustina Baccin e João Azenha Jr. por sua valiosa contribuição e pelos direcionamentos sugeridos em meu exame de qualificação. Ao Departamento de Lingüística da Universidade do Texas em Austin, por me receber como bolsista de doutorado, e aos professores do Computational Linguistics Lab – Prof. Dr. Jason Baldridge, pela orientação de meu projeto de estágio sanduíche, e à Profa. Dra. Katrin Erk, pelas aulas sobre Processamento da Linguagem Natural, programação em Python e Estatística aplicada à Lingüística de Corpus.

Agradeço à Coordenadoria do Programa de Estudos Lingüísticos e Literários em Inglês do Departamento de Letras Modernas, que auxiliou no provimento das condições necessárias para a realização de minha pesquisa. À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pelo auxílio financeiro que me foi concedido durante parte do doutorado no Brasil e para o estágio de seis meses que fiz na Universidade do Texas em Austin. Ao Conselho Nacional de Pesquisa (CNPq), pelo auxílio financeiro concedido nos meses finais do doutorado. Aos funcionários do Departamento de Letras Modernas – em especial à Edite, que ouviu com paciência minhas reclamações, e sempre fez o que estava a seu alcance para me auxiliar. À Comissão de Cooperação Internacional, especialmente à simpaticíssima Rosângela Duarte Vicente (cuja eficiência eu já conhecia dos tempos da Especialização em Tradução), que foi quem primeiro me ajudou na empreitada de estagiar no exterior. Ao João Roberto Patrinhani, meu excelente ex-aluno de inglês, funcionário da Pró-Reitoria

de Pós-graduação da USP, que me guiou pelos áridos caminhos da burocracia – preencher e submeter a infindável papelada, em meio à greve, invasão de reitoria, etc.

Meus agradecimentos a todos os colegas de orientação e de Grupo Comet – ao Alfredo, à queridíssíssima Alva, Andréa, Carla Nejjjjjm, Ca(r)mem, Cris, Gisele, Guifromm, Helmara, Josi, à Lourdinha-docinho-de-coco, às divertidíssimas “Lús” (só vocês pra me fazerem rir nos momentos mais difíceis), às inseparáveis e animadíssimas “Marlenes”, Rozane, Sabrina, Sérgio (ai, será que esqueci de alguém?... espero que não!) – e aos alunos e orientandos do Prof. Tony na PUC, pelas proveitosas discussões sobre a Lingüística de Corpus e suas múltiplas aplicações. Aos alunos das duas últimas turmas do já saudoso Curso de Especialização em Tradução da USP, e também aos alunos das turmas de Tradução do Unibero, de inglês conversação e instrumental da Reitora da USP e de inglês técnico da Faculdade Módulo, que tiveram aula comigo – aprendi muito ao conviver e trabalhar com todos vocês.

A todos que moram em meu coração – sem seu amor e apoio teria sido muito mais difícil chegar ao fim dessa tortuosa jornada. À minha família, que sempre me apoiou em meus estudos, mas também me ensinou que a verdadeira realização pessoal advém da satisfação de fazer o que se gosta. A todos os meus amigos queridos, pela paciência que têm sempre comigo, em minhas idas e vindas – vocês são muito importantes pra mim. Não ousa, a essa altura do campeonato, e ante ao atual estado de minhas faculdades mentais, ter a pretensão de enumerar seus lindos nomes (Pop Gomes, gracias por traduzir o “resumen”!), mas cada um de vocês que teve a paciência de ler essas linhas iniciais saberá que meu agradecimento é também para você, do fundo do coração. À Carol que, como sempre, me pegou pela mão e conduziu de volta à luz nos momentos mais difíceis, me aconselhando e escutando minhas infindáveis lamentações – você sempre foi e sempre será meu modelo de pesquisadora e, especialmente, de pessoa.

Enfim, agradeço a todos aqueles que direta ou indiretamente me apoiaram e colaboraram com a minha pesquisa – inclusive meus dois gatos fofos, Juca e Ísis, que agora são internacionais!

SUMÁRIO

Resumo.....	i
<i>Abstract</i>	ii
<i>Resumen</i>	iii
Índice de figuras.....	iv
Índice de tabelas.....	vi
Lista de abreviações.....	vii
Índice geral.....	viii
Introdução.....	2
Capítulo I – Tradução técnica: teoria(s) e algumas questões práticas da interface com a Terminologia.....	14
Capítulo II – Tradução culinária no Brasil: mercado de trabalho, materiais de apoio e suas lacunas.....	68
Capítulo III – Lingüística de Corpus e dicionários: mais uma metodologia, ou uma nova forma de observar textos especializados?.....	150
Capítulo IV – Compilando Unidades de Tradução Especializadas em receitas culinárias via Lingüística de Corpus.....	219
Capítulo V – Proposta de um Dicionário <i>Online</i> de Culinária voltado para o tradutor.....	323
Conclusão.....	340
Referências Bibliográficas.....	345
Anexos.....	368

RESUMO

TEIXEIRA, E. D. **A Lingüística de Corpus a serviço do tradutor: proposta de um dicionário de Culinária voltado para a produção textual**. 2008. 400 f. Tese (Doutorado) – Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2008.

Os dicionários sempre foram – e ainda são – uma das principais ferramentas da tarefa tradutória. No entanto, a terminografia parece não ter se beneficiado ainda de forma sistemática, pelo menos no Brasil, da estreita relação entre dicionários técnicos e esse público-alvo específico e cada vez mais expressivo: os tradutores técnicos. Na área da Culinária, por exemplo, cuja demanda por traduções tem crescido regularmente no país, os dicionários disponíveis no par de línguas inglês-português, ainda que possam contribuir para a compreensão do texto original, não dão qualquer informação sobre como os termos são, de fato, usados em textos reais na área, ou seja, não auxiliam o tradutor numa etapa fundamental da atividade tradutória: a produção textual na língua de chegada. Nossa tese é a de que um dicionário que procure atender as necessidades de produção textual do tradutor deve concentrar-se nos aspectos que caracterizam o texto técnico do ponto de vista da tradução, isto é, deve descrever e propor equivalentes ou soluções tradutórias para as Unidades de Tradução Especializadas (UTES) presentes nesses textos, com as quais o tradutor da área se depara freqüentemente em sua prática, sejam elas terminológicas ou não. A Lingüística de Corpus (L.C.), abordagem empirista que vê a língua como um sistema probabilístico, tem se dedicado à identificação de padrões léxico-gramaticais recorrentes na linguagem por meio da observação de textos autênticos organizados sob a forma de *corpora* eletrônicos. É, portanto, a área de estudos que julgamos fornecer os subsídios teóricos e metodológicos mais adequados para compilar as UTES a partir de textos reais. Para abrigar essas unidades, apresentamos uma proposta de dicionário *online* bidirecional inglês-português voltado para o tradutor da área técnica da Culinária. As etapas seguidas no desenvolvimento deste trabalho estão organizadas em seis capítulos. O primeiro trata de aspectos teórico-práticos da tradução técnica e discute o papel da terminologia na prática tradutória. O segundo examina as especificidades do trabalho do tradutor da área da Culinária no Brasil e caracteriza a “receita culinária”, foco de nosso estudo, como gênero e tipologia textual. No terceiro capítulo, as bases teóricas e metodológicas da L.C. são apresentadas, bem como os critérios de coleta do *corpus* que servirá de base para a identificação das UTES. O Capítulo IV descreve a etapa de exploração desse *corpus*: primeiramente, apresentamos o estudo em que é feito um levantamento manual de padrões léxico-gramaticais usando o programa WordSmith Tools; em seguida, descrevemos uma metodologia de extração semi-automática de UTES no *corpus*. O Capítulo V apresenta nossa proposta do Dicionário *Online* de Culinária bidirecional para Tradutores, descrevendo sua macro- e microestrutura. O Capítulo VI apresenta nossas considerações finais. Os resultados obtidos em nosso trabalho demonstram que a L.C., se usada não apenas como metodologia, mas como abordagem teórica na exploração de *corpora* especializados, permite elaborar dicionários mais úteis e confiáveis para o tradutor, pois leva em consideração quaisquer padrões associativos entre palavras cuja probabilidade de ocorrer em textos representativos da área seja alta – fato que justifica plenamente a inclusão desses padrões num dicionário voltado para o tradutor-produtor-de-textos.

PALAVRAS-CHAVE: Lingüística de Corpus. Tradução técnica. Terminologia. Dicionários bilíngües. Culinária.

ABSTRACT

TEIXEIRA, E.D. **Corpus Linguistics at the translator's service: proposal of an online dictionary of Culinary aiming at text production.** 2008. 400 p. PhD dissertation – Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2008.

Dictionaries have always been – and still are – one of the main tools for the translator's task. Nevertheless, terminography does not seem to have systematically benefited, at least in Brazil, from the close relation between technical dictionaries and this increasingly significant and particular target audience: technical translators. In the field of Culinary, for instance, which has witnessed a growing demand for translations in Brazil, much as the dictionaries available for the English-Portuguese language pair may contribute to the understanding of the source text, they still do not provide any information on how the terms are actually used in real texts. In other words, they do not help the translator in a crucial step of the translation activity: text production in the target language. The thesis advanced here is that a dictionary which seeks to meet the needs of a translator's text production should focus on the aspects which characterize technical texts from the point of view of translation itself, that is, it should describe and propose translation equivalents or suggestions for the Specialized Translation Units (STUs) occurring in these texts, which the translator in the area often comes across in her/his practice, whether they are terminological or not. Corpus Linguistics (CL), an empirical approach which regards language as a probabilistic system, has devoted itself to the identification of recurring lexico-grammatical patterns in language by observing authentic texts organized as electronic *corpora*. It is, therefore, the field of studies we deem capable of providing the most adequate theoretical and methodological support to extract the STUs from real texts. In order to embrace these units, a proposal of an online bidirectional English-Portuguese dictionary is presented, which is aimed at the technical Culinary translator. The steps followed in the development of this study were organized in six chapters. The first deals with the theoretical and practical aspects of technical translation and discusses the role of terminology in translation practice. The second examines the specificities of the Culinary translator's job in Brazil and characterizes the "culinary recipe", the focus of this study, in terms of text genre and typology. In the third chapter, the theoretical and methodological foundations of CL are presented, as well as the criteria used in the compilation of the corpus to be used for the identification of the STUs. Chapter IV describes the exploration of this *corpus*: first, we present the study in which a manual search of lexico-grammatical patterns using the WordSmith Tools program is carried out; next, a methodology for semi-automatic extraction of STUs in the *corpus* is described. Chapter V presents our proposal for a bidirectional online Culinary Dictionary for Translators, describing its macro- and microstructure. Chapter VI contains the final considerations. The results obtained in this study bear witness to the fact that CL, if used not only as a methodology, but as a theoretical approach in the investigation of specialized *corpora*, enables the production of more useful and trustworthy dictionaries for the translator, for it takes into account any association patterns between words with a probability of occurrence in representative texts in the field – a fact which strongly supports the inclusion of these patterns in a dictionary aimed at the translator-producer-of-texts.

KEYWORDS: Corpus Linguistics. Technical Translation. Terminology. Bilingual dictionaries. Culinary.

RESUMEN

TEIXEIRA, E. D. **La Lingüística de Corpus ayudando al traductor: propuesta de un diccionario de Culinaria dirigido a la producción textual**. 2008. 400 p. Tesis (Doctorado) – Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2008.

Los diccionarios siempre han sido – y todavía lo son – una de las principales herramientas para la práctica traductora. Sin embargo, parece que la terminología en Brasil no se ha beneficiado aún de forma sistemática de la estricta relación entre diccionarios técnicos y ese grupo de usuarios específico y cada vez más expresivo: los traductores técnicos. En el área de la Culinaria, por ejemplo, cuya procura por traducciones aumenta regularmente en el país, los diccionarios bidireccionales disponibles de inglés y portugués, aunque puedan contribuir a la comprensión del texto original, no proveen información sobre como realmente son los términos utilizados en textos culinarios, es decir, no auxilian el traductor en una etapa fundamental de la actividad traductora: la producción textual en la lengua de llegada. Nuestra tesis es que un diccionario que busque atender las necesidades de producción textual del traductor debe concentrarse en aspectos que caracterizan el texto técnico desde la perspectiva de la traducción, o sea, debe describir y proponer equivalentes o soluciones de traducción para las Unidades de Traducción Especializadas (UTES) presentes en estos textos, con las que el traductor del área se encuentra a menudo en su práctica, sean terminológicas o no. La Lingüística de Corpus (L.C.), aproximación empirista que considera la lengua un sistema probabilístico, tiene se dedicado a la identificación de patrones léxico-gramaticales recurrentes en el lenguaje a través de la observación de textos auténticos organizados como forma de *corpora* electrónicos. Luego, la L.C. es el área de estudio que juzgamos proveer los subsidios teóricos y metodológicos más adecuados para compilar las UTES desde textos reales. Para acoger esas unidades, presentamos una propuesta de diccionario *online* bidireccional inglés-portugués dirigido al traductor del área técnica de Culinaria. Las etapas seguidas para el desarrollo de este trabajo están organizadas en seis capítulos. El primero trata de aspectos teórico-prácticos de la traducción técnica y discute el papel de la terminología como práctica traductora. El segundo examina las especificidades del oficio de traductor del área de Culinaria en Brasil, y caracteriza la “receta culinaria”, foco de nuestro estudio, como género y tipología textual. En el tercer capítulo, las bases teóricas y metodológicas de L.C. son presentadas, también los criterios de colecta del *corpus*, que se prestará para la identificación de las UTES. El Capítulo IV describe la etapa de exploración de ese *corpus*: primeramente presentamos el estudio con un inventario manual de patrones léxico-gramaticales utilizando el programa WordSmith Tools; a continuación describimos una metodología de extracción semiautomática de UTES en el *corpus*. El Capítulo V presenta nuestra propuesta para el Diccionario Online de Culinaria bidireccional para Traductores, describiendo su macro y microestructura. El Capítulo VI presenta nuestras consideraciones finales. Los resultados obtenidos en nuestra investigación demuestran que la L.C., si no es utilizada sólo como metodología, sino como abordaje teórica en la explotación de *corpora* especializados, permite elaborar diccionarios más útiles y confiables para el traductor porque considera cualquier patrón asociativo entre palabras cuya probabilidad de ocurrir en textos representativos del área sea alta – lo que justifica totalmente la inclusión de estos patrones en un diccionario dirigido al traductor-productor de textos.

PALABRAS-CLAVE: Lingüística de Corpus. Traducción técnica. Terminología. Diccionarios bilingües. Culinaria.

Índice das figuras

Fig.	Título	pág.
1	Estudos da Tradução para Holmes (baseado em Baker 1999: 278)	17
2	Categorização dos Procedimentos Técnicos da Tradução	45
3	Quadro-resumo baseado em Gomes e Barbosa (2004)	76
4	Informações nutricionais do creme de leite Nestlé® em português e inglês	87
5	Lista de ingredientes – pacote de granola	87
6	Instruções de preparo e receitas – sopa Campbells®	88
7	Exemplo de menu em inglês – restaurante Tordesilhas, SP	92
8	Exemplo de receita com fotos do passo a passo e do prato final	98
9	Exemplo de inconsistência entre receita a foto	110
10	Exemplo de alteração de ingrediente compatível com a foto	111
11	Níveis de especialização do ponto de vista textual	116
12	Exemplo de receita com verbo na 1a. pessoa do singular	121
13	Exemplo de verbete "semântico" em Teixeira e Tagnin (no prelo)	134
14	Verbetes "lemon" em Teixeira e Tagnin (no prelo)	135
15	Configurações para usar o KWIC Google	143
16	Tela de resultado do KWIC Google	144
17	Tela de resultados do concordanciador do CorTec	145
18	Linhas de concordância ordenadas pelas palavras à direita	146
19	Exemplo de cabeçalho preenchido	163
20	Resultados da ferramenta WordList (listador de palavras)	171
21	Configuração para compilar wordlists com mais de 1 palavra	172
22	Resultados do Wordlist com clusters de 3 palavras	172
23	Resultado da ferramenta KeyWords (listador de palavras-chave)	174
24	Resultado de busca para "pimentão" no concordanciador (Concord)	175
25	Resultado de "pimenta" ordenado pelos colocados à direita	175
26	Resultado de "pimenta" ordenado pelos colocados à esquerda	176
27	Funções acessadas pela barra de ferramentas do Concord	177
28	Tabela de colocados de <i>picad*</i> gerada pelo Concord	178
29	Tabela de padrões de colocados (<i>patterns</i>)	179
30	Ajuste dos parâmetros para a extração de <i>clusters</i>	180
31	Lista de <i>clusters</i> gerada pelo Concord	181
32	Exemplo de tradução não-natural	186
33	Linhas de concordância para a palavra "flour" seguida de "egg"	190
34	Primeiras palavras-chave dos <i>corpora</i> de Culinária de português e inglês	191
35	Exemplo de contexto expandido e acesso ao arquivo original via Concord	193
36	Tela de resultado do VoTec para o verbete "motherboard"	197
37	Verbetes "label" em Dancette e Réthoré (2000)	199
38	Verbetes "label" no <i>DAD enLigne</i> (parte superior da tela)	201

39	Verbetes “label” no <i>DAD enLigne</i> (parte inferior da tela)	202
40	Busca livre de “optical” no <i>DAD enLigne</i>	203
41	Como acessar o utilitário <i>Text Converter</i> do WST	210
42	Tela principal do utilitário <i>Text Converter</i>	211
43	Exemplo de receita etiquetada	212
44	Linhas de concordância para a palavra de busca “*flour*”	223
45	Glossário proveniente da memória de tradução	239
46	Transformando colunas do Excel em texto	241
47	Transformando marcas de parágrafo em ponto final	241
48	Arquivo em .txt pronto para ser enviado para o etiquetador	241
49	Etiquetador do CEPRIL	242
50	Resultado do etiquetador colado em planilha do Excel	243
51	Aparência do texto durante a formatação dos resultados do Tree-tagger	244
52	Tabela com as unidades, categorias e n°. de elementos	245
53	Uso da ferramenta <i>Sort & Filter</i> para localizar erros do etiquetador	247
54	Relação entre número de elementos e categoria gramatical	257
55	Lista de palavras (Wordlist) do corpus de culinária – português	259
56	Lista de palavras-chave (keywords) do corpus de português	260
57	Janela em que se escolhe o <i>p</i> a ser usado na <i>Keyword List</i>	260
58	Emulador de <i>Unix Cygwin</i> rodando em ambiente <i>Windows</i>	261
59	Palavras-chave em .txt	262
60	UTES itemizadas	262
61	Rotina em Shell para procurar palavras-chave	263
62	UTES com .* substituindo a parte variável	268
63	Aparência da tabela final	269
64	Linhas de concordância para “abr*” + “enfarinhada”	271
65	Linhas de concordância para “até que... evapor*”	273
66	Linhas de concordância para “até que”	275
67	Linhas de concordância para “até que... ao”	276
68	Tela do editor IDLE e <i>prompt</i> do programa <i>Python</i>	284
69	Aparência do arquivo gerado pelo programa	289
70	Aparência da nova tabela Excel com os resultados	292
71a	Total geral de padrões recuperados por categoria	295
71b	Proporção de formas distintas (<i>types</i>) ocorrendo 3 ou mais vezes	295
72	Usando a função “Random” do Excel	297
73	Página inicial do Dicionário Online de Culinária para Tradutores	326
74	Consulta do DOC-Trad por palavra-chave	327
75	Aparência da árvore hiperbólica da Culinária	328
76	<i>Links</i> para outros <i>sites</i> relacionados à Culinária ou à pesquisa na <i>Web</i>	330
77	Exemplo de verbete português-inglês	338
78	Exemplo de verbete inglês-português	339

Índice das tabelas

Tab.	Título	pág.
1	Terminologia alimentar de acordo com Baccin (2003)	114
2	Exemplos de hipercorreção na revisão de receitas por editoras	118
3	Termos relacionados a “pepper” no <i>Tradutor Gastronômico</i>	136
4	Critérios iniciais usados na coleta do corpus de Culinária	209
5	Padrões nominais de “flour*”	224
6	Padrões verbais de “flour*”	226
7	Número de ocorrências por categoria e porcentagem de erros	247
8	Substantivos e sintagmas iniciados por substantivo	248
9	Verbos e sintagmas iniciados por verbo	251
10	Adjetivos e sintagmas iniciados por adjetivos	253
11	Sintagmas iniciados por preposições	253
12	Advérbios e UTEs iniciadas por advérbios	254
13	UTES iniciadas por conjunções	254
14	UTES iniciadas por outras categorias	255
15	Faixas de frequência das UTEs	269
16	Dados quantitativos gerais sobre a extração por padrões	293
17	Proporção entre formas distintas (<i>types</i>) e frequência (<i>tokens</i>)	296
18	Quadro-resumo dos resultados da análise das amostras aleatórias	298
19	Campos da ficha de recolha de dados	331

Lista de abreviaturas

arq.	arquivo(s)
CBTS	Corpus-based Translation Studies
DOC-Trad	Dicionário Online de Culinária para Tradutores
Fig.	Figura
L.C.	Linguística de Corpus
LCh	Língua de Chegada
LP	Língua de Partida
oc.	Ocorrência(s)
PLN	Processamento da Linguagem Natural
Tab.	Tabela
TCT	Teoria Comunicativa da Terminologia
TGT	Teoria Geral da Terminologia
TST	Teoria Sociocognitiva da Terminologia
UTE	Unidade de Tradução Especializada
WST	Word Smith Tools

ÍNDICE

Introdução	2
Capítulo I – Tradução técnica: teoria(s) e algumas questões práticas da interface com a Terminologia	14
1.1 Tradução técnica: prática e reflexão sobre a prática	14
1.2 A Tradução técnica no âmbito das teorias da tradução	19
1.2.1 Texto técnico e cultura	21
1.2.2 Texto técnico e usabilidade	27
1.2.3 Texto técnico e tipologia textual	32
1.2.4 “Tradução técnica” neste trabalho	34
1.2.5 Alguns conceitos básicos da Tradução	36
1.2.5.1 Unidade de tradução	36
1.2.5.2 Equivalência e estratégia de tradução	41
1.2.5.3 Naturalidade	49
1.3 Tradução técnica e dicionários	51
1.3.1 Tradução técnica e terminologia: caminhos entrecruzados	53
1.3.2 Wüster revisitado: uma nova visão do texto técnico	56
1.3.3 Terminologia Textual	62
Capítulo II – Tradução culinária no Brasil: mercado de trabalho, materiais de apoio e suas lacunas	68
2.1 Globalização e o <i>boom</i> da Culinária	69
2.1.1 A Culinária no mercado editorial	74
2.1.2 A Culinária no mercado de tradução brasileiro	79
2.2 Breve mapeamento da Culinária	84
2.2.1 Embalagens e rótulos de alimentos e bebidas	85
2.2.1.1 Informações nutricionais	86
2.2.1.2 Lista de ingredientes	86
2.2.1.3 Instruções de preparo, sugestão de consumo e receitas	87
2.2.1.4 Nome do produto, descritivo, foto e outros elementos	88
2.2.1.5 Bebidas alcoólicas: um caso a parte	88

2.2.2 Cardápios de restaurantes.....	90
2.2.3 Interpretação para turismo gastronômico.....	92
2.2.4 Legendagem de material audiovisual sobre culinária.....	93
2.2.5 Tradução de material multimídia.....	93
2.2.6 Tradução de livros e outros materiais impressos.....	93
2.3 Receita: o texto culinário por excelência.....	94
2.3.1 Receitas quanto ao meio de divulgação.....	95
2.3.2 Receitas quanto à organização e à aparência.....	97
2.3.3 Componentes estruturais e lingüísticos da receita.....	98
2.3.3.1 Elementos obrigatórios.....	100
2.3.3.2 Elementos opcionais.....	107
2.3.4 Receitas quanto ao grau de especialização.....	113
2.3.5 Receitas e estilo.....	117
2.4 Tradução culinária no par de línguas inglês – português.....	122
2.4.1 Problemas de não-equivalência e equivalência parcial em Culinária.....	123
2.4.2 Ferramentas de apoio ao tradutor da área.....	130
2.4.2.1 Dicionários.....	130
2.4.2.2 Outros tipos de material impresso.....	139
2.4.2.3 Internet.....	141
2.4.2.4 Especialista.....	146
2.4.3 O que o tradutor de Culinária precisa.....	147
 Capítulo III – Lingüística de Corpus e dicionários: mais uma metodologia, ou uma nova forma de observar textos especializados?	150
3.1 Lingüística de Corpus: brevíssima introdução.....	151
3.2 Lingüística de Corpus como abordagem e como metodologia.....	154
3.3 Princípios basilares da Lingüística de Corpus.....	158
3.3.1 Corpus.....	158
3.3.1.1 Tipos de corpus.....	161
3.3.2 Representatividade.....	164
3.3.3 Padronização.....	166
3.3.4 Ferramentas de exploração de corpora.....	169
3.3.4.1 Listadores de palavras.....	170
3.3.4.2 Listadores de palavras-chave.....	172
3.3.4.3 Concordanciadores.....	174
3.3.4.4 Tabela de colocados.....	177
3.3.4.5 Lista de <i>clusters</i>	180
3.4 Lingüística de Corpus e Tradução.....	181
3.4.1 Naturalidade e equivalência da perspectiva da L.C.....	185
3.4.2 Lingüística de Corpus a serviço da tradução.....	187

3.4.3	Linguística de Corpus e dicionários técnicos.....	193
3.5	Compilação do corpus comparável de Culinária.....	206
3.5.1	Planejamento do corpus.....	206
3.5.2	Coleta, limpeza e organização dos textos do corpus.....	209
3.5.3	Alguns dados gerais sobre o corpus.....	211
3.5.3.1	Subcorpus de inglês.....	211
3.5.3.2	Subcorpus de português.....	214
3.5.4	Considerações finais sobre a compilação do corpus.....	217

Capítulo IV – Compilando Unidades de Tradução Especializadas em receitas culinárias via Linguística de Corpus.....		219
4.1	Estudo da padronização de uma palavra-chave do corpus.....	221
4.1.1	O ponto de partida.....	222
4.1.2	Padrões de “flour” identificados no subcorpus de inglês.....	224
4.1.2.1	Padrões nominais.....	224
4.1.2.2	Padrões verbais.....	225
4.1.2.3	Considerações preliminares sobre os dados levantados.....	228
4.1.3	Procurando equivalentes para os padrões em dicionários.....	229
4.1.3.1	Dicionários bi- e multilíngües.....	229
4.1.3.2	Dicionários monolíngües.....	232
4.1.4	Cotejando os dados.....	233
4.1.5	Considerações finais sobre o estudo.....	236
4.2	Estudo das Unidades de Tradução provenientes da memória de tradução.....	237
4.2.1	O ponto de partida.....	237
4.2.2	Associação entre categorias gramaticais.....	239
4.2.2.1	Procedimento de etiquetagem.....	240
4.2.2.2	Erros do etiquetador Tree-Tagger.....	245
4.2.2.3	Distribuição das UTEs por categoria morfo sintática.....	249
4.2.2.4	Análise dos dados e algumas considerações preliminares.....	255
4.2.3	UTES e palavras-chave.....	258
4.2.3.1	Verificando a presença de palavras-chave nas UTEs.....	261
4.2.3.2	Análise dos dados e algumas considerações preliminares.....	264
4.2.4	Frequência das UTEs no corpus de português.....	267
4.2.4.1	Verificando a frequência das UTEs no corpus total.....	267
4.2.4.2	Análise dos dados e algumas considerações preliminares.....	269
4.2.5	Frequência dos padrões morfo sintáticos no corpus total.....	277
4.2.5.1	Melhorando a etiquetagem do corpus.....	278
4.2.5.2	Extração das frequências por categoria inicial da UTE.....	287
4.2.5.3	Frequência dos padrões no corpus de Culinária total.....	293
4.2.5.4	Análise da precisão em amostras aleatórias.....	296

4.2.5.5 Resumo das estratégias para melhorar a extração.....	317
4.3 Considerações finais sobre a extração semi-automática de UTEs.....	318

Capítulo V – Proposta de um Dicionário *Online* de Culinária voltado para o tradutor.....

5.1 Porque um dicionário de Culinária <i>online</i>	324
5.1.1 Macroestrutura: a interface do Dicionário <i>Online</i> de Culinária.....	325
5.1.2 A ficha para registro dos dados.....	331
5.1.3 Microestrutura: o verbete.....	333
5.1.3.1 Entrada.....	333
5.1.3.2 Paradigma informacional.....	334
5.1.3.3 Paradigma pragmático.....	335
5.1.3.4 Paradigma de forma equivalente.....	336
5.1.4 Exemplos de verbete.....	337
5.1.4.1 Direção português-inglês.....	338
5.1.4.2 Direção inglês-português.....	339

Conclusão.....

Referências Bibliográficas.....

Anexo I – Dados do Corpus Comparável de Culinária.....	366
Anexo II – Palavras-chave do subcorpus de Culinária – português.....	369
Anexo III – Palavras-chave do subcorpus de Culinária – inglês.....	370
Anexo IV – Equivalentes para os padrões de “flour*” em dicionários.....	371
Anexo V – Dados sobre a metodologia semi-automática de extração.....	380
Anexo VI – 100 UTEs mais recorrentes nas categorias (ordem alfabética).....	395
Anexo VII – Exemplo de visualização inicial de verbete no DOC-Trad.....	399

*It is not the words which tell you the meaning of the phrase,
but the phrase which tells you
the meaning of the individual words within it.*

Michael Stubbs (2001: 14)

*Quem acha que a Lingüística de Corpus não passa
de uma ferramenta que nada ou pouco contribui
para repensar os fundamentos da Lingüística, tal qual
ela se desenvolveu, digamos, nos últimos 50 anos,
está prestes a ser saudado com uma bela surpresa.*

Kanavillil Rajagopalan (2007: 32)

This document was created with Win2PDF available at <http://www.daneprairie.com>.
The unregistered version of Win2PDF is for evaluation or non-commercial use only.

Introdução

Los glosarios elaborados por terminógrafos no suelen satisfacer las necesidades de los traductores porque no parten del análisis de las necesidades reales que plantea la traducción. Por ello, clamamos para que el colectivo de la profesión traductora entre con mayor profusión en la elaboración de recursos terminológicos que respondan más coherente y adecuadamente a las necesidades reales de la traducción
(Cabré, 1999a:12)

a organização de instrumentos de referência temática não monolíngües, tarefa que os especialistas em Terminologia tomam a si, só pode ser adequadamente realizada com o concurso de tradutores
(Krieger e Finatto, 2004: 69)

Tradução Técnica e Terminologia são áreas de estudo co-relacionadas, reconhecem vários autores de ambas as áreas (Aubert, 1992 e 1996; Azenha Jr., 1999; Cabré, 1999 e 2002; Krieger e Finatto, 2004; Barros, 2004 e 2006, entre outros). No entanto, apesar de todos os avanços teóricos ocorridos no mundo nos últimos anos nesses dois campos de pesquisa, a prática terminográfica em si parece não ter se beneficiado ainda, de forma sistemática – pelo menos no Brasil – dessa estreita relação que têm os dicionários técnicos com um público-alvo específico e cada vez mais expressivo: o tradutor.

Uma boa parte dos dicionários bilíngües especializados disponíveis atualmente em português brasileiro, salvo algumas exceções¹, consiste de meras listas de termos, muitos deles monovocabulares, por vezes recopiladas a partir de outras listas semelhantes, baseadas em informações desatualizadas e de confiabilidade duvidosa (Teixeira, 2003 e 2004). E essa falta de rigor metodológico não é exclusividade das obras terminográficas – pode ser observada inclusive em dicionários de língua geral publicados em território nacional (Biderman, 2004) que, muitas vezes, são até mesmo usados como fonte de informação “terminológica” na compilação de dicionários especializados.

Há áreas técnicas, é certo, que contam com maior número de obras, algumas muito bem elaboradas – no geral, campos do saber que desfrutam de grande prestígio social e/ou econômico, que lhes garante maior visibilidade e consideração (como é o caso, por exemplo, do Direito, da Medicina, da Economia e suas respectivas subáreas). Mas, no geral, dicionários técnicos são escassos, quando não diminutos e incompletos. E a situação piora exponencialmente quando se trata de obras bilíngües, ainda que no Brasil o par de línguas português-inglês seja um dos que mais tem rendido publicações.

Dicionários sempre foram – e ainda são – uma das principais ferramentas de trabalho do tradutor, e na tradução técnica não poderia ser diferente. No entanto, ao passo que a Lexicologia (campo de estudo tradicional da Lingüística há muito mais tempo que a Terminologia) evoluiu bastante no sentido de diversificar os tipos de obra produzidos e as metodologias de compilação de forma a atender diferentes necessidades de diferentes públicos, o mesmo parece não ter ocorrido, na prática, com a Terminologia.

Se há hoje uma série de “dicionários especiais” de língua geral (Barros, 2006: 23), voltados para diferentes grupos de usuários, ou para diferentes necessidades dos usuários desses grupos – como dicionários de expressões idiomáticas, de sinônimos, de antônimos, de verbos, de gírias, de provérbios, dicionários escolares, vocabulários dedicados a certos autores consagrados da literatura e até mesmo dicionários de uso baseados em *corpora* eletrônicos, entre tantos outros –, em Terminologia, o público-alvo parece ser sempre o mesmo: o especialista da área. Ou, mais especificamente, o especialista em formação.

¹ Vide, por exemplo, os dicionários produzidos pelo Grupo Termisul (<http://www6.ufrgs.br/termisul/>), pelo Grupo TextQuim (<http://www6.ufrgs.br/textquim/>) ou Perroti-Garcia e Rebechi (2007).

O tradutor, talvez por questões mercadológicas, costuma ser evocado na contracapa, na orelha ou na introdução de dicionários técnicos multilíngües, mas percebe logo a falácia ao tentar usar a obra recém-adquirida em sua prática diária, especialmente nos trabalhos de versão. De fato, na maioria desses dicionários, o autor parece ter se lembrado muito pouco do tradutor, ou até mesmo ter ignorado suas necessidades quando da compilação da obra. Onde estão os exemplos de uso? Onde se explica o grau de equivalência entre os termos? Qual dos dois ou mais equivalentes oferecidos (nos raros casos em que há mais de um) usar no texto de chegada? Que preposição usar com que verbo, que adjetivo com que substantivo, onde estão os advérbios?

Essas e tantas outras perguntas torturam a mente do tradutor, especialmente o inexperiente, ou aquele que está começando a trabalhar numa área especializada nova, quando tenta usar os poucos dicionários que encontra (quando os encontra), sejam eles impressos ou eletrônicos, para produzir um texto na língua de chegada que, obviamente, seja preciso – uma das características mais valorizadas na tradução técnica –, mas também natural. O que falta, de fato, nessas obras de referência para que sejam mais úteis e mais confiáveis para o tradutor? Por que essa demanda, já tão antiga, ainda não foi atendida a contento pela prática terminográfica?

Para tentar responder essas perguntas, e compreender melhor as necessidades desse tradutor técnico em sua prática diária – profissão que, ao menos no Brasil, sofre de um desprestígio histórico – é que realizei o presente trabalho. Meu principal objetivo é propor um modelo de dicionário *online* bidirecional português-inglês voltado para as necessidades desse público-alvo específico. Para tanto, precisei rever, de perto, as relações entre Tradução Técnica e Terminologia, o que fiz à luz da Linguística de Corpus, pelos motivos que passo a explicar agora.

A Terminologia, desde sua gênese, tem no termo e na definição seus objetos de estudo privilegiados. Só muito recentemente a fraseologia terminológica – estrutura representativa de um nódulo conceitual que inclui um termo em sua composição (Krieger e Finatto, 2004: 85) – começou a ser estudada e descrita em dicionários especializados (Blais, 1993; Gouadec, 1994; Desmet, 1995-6; Bevilacqua, 1996, 1999; Cabré, Lorente e Estopá, 1996; Estopá, 1999, entre outros). Segundo Bevilacqua (2001: 106) e Krieger e Finatto (2004: 86), uma das maiores motivações para isso

advém do próprio mercado de tradução técnica, em franca expansão desde a Segunda Guerra Mundial.

Sabe-se que a tradução técnica ocupa a maior fatia do mercado de traduções atualmente, seja no panorama nacional, seja no mundial – 90% de todos os trabalhos de tradução feitos no mundo, de acordo com Kingscott (2002: 247), 70% da demanda por traduções no Brasil, segundo Finatto (2001: 56). No entanto, conforme pontuam Byrne (2006: 2-7) e Korning Zethsen (1997, 1999), entre outros, há muitas crenças errôneas a respeito da tradução técnica que não só prejudicam seu adequado exercício mas, a meu ver, impedem um melhor aproveitamento, por parte da terminografia, das características dessa prática e de suas necessidades no que diz respeito, especialmente, à produção textual na Língua de Chegada (doravante LCh).

Para Korning Zethsen (1999), essas crenças errôneas têm em comum uma concepção: o texto técnico se reduz à terminologia que veicula. Serve para transferir um conhecimento técnico-científico (que existe independentemente da linguagem, do texto e da cultura) e, idealmente, é desprovido de qualquer forma de estilo ou identidade lingüístico-cultural.

O tipo de texto geralmente privilegiado por algumas linhas teóricas para compor os *corpora* usados para o levantamento de termos e definições, por exemplo, reflete uma preocupação com esse conteúdo conceptual. Textos de natureza mais didática ou divulgativa, que contêm mais contextos definitórios e explicativos, são privilegiados, muitas vezes, em detrimento dos altamente especializados, escritos por especialistas para serem lidos por especialistas (em geral, o tipo de texto mais traduzido / vertido), por serem mais úteis ao terminólogo na tarefa de identificar e definir os conceitos de uma dada área. Mas em que medida são representativos do universo de discurso da área em questão e, mais especificamente, do tipo de texto com o qual o tradutor precisa lidar no dia-a-dia de sua profissão?

A meu ver, tal escolha metodológica pode ser um indício de que a prática terminográfica, de certa forma, tem deixado a materialidade lingüístico-textual do conhecimento especializado em segundo plano, ainda que os motivos pareçam perfeitamente justificáveis pelos modelos teóricos adotados (foco no conceito), e a julgar pelo tipo de dicionário técnico resultante e o público-alvo que privilegia (o especialista-aprendiz). Em outras palavras: apesar de a comunicação entre especialistas ser levada em conta pelos modelos teóricos mais atuais, termo, conceito

e definição são o objeto central de estudo da Terminologia e principal alvo da compilação e descrição em dicionários especializados. O estudo do texto técnico *per se*, do entorno cultural subjacente às suas condições de produção e recepção, bem como do “habitat lingüístico” dos termos (seu co-texto) parecem não ter se consolidado, ainda, como indispensáveis na compilação de obras terminográficas bilíngües, apesar de serem da maior importância para o tradutor na produção textual na língua de chegada.

A abordagem adotada neste trabalho parte do pressuposto de que, na tradução técnica, o correto uso da terminologia da área não garante, isoladamente, que uma tradução seja bem sucedida (Azenha Jr., 1999; Byrne, 2006; Korning Zethsen, 1999, entre outros). Newmark (1988: 160), a propósito, chegou a afirmar que a terminologia representaria algo entre 5 e 10% do texto técnico: “The rest is ‘language’...” [O resto é ‘língua(gem)’...]. Assim, uma tradução bem sucedida dependeria, na maioria das vezes, da habilidade do tradutor em produzir um texto natural, que possa ser “lido como original” (Fishback, 1998: 2, *apud* Byrne, 2006: 4, 15). Para isso, o tradutor necessita não só ter conhecimento dos conceitos e termos da área, mas ter ótima redação, e levar em conta as características do gênero e tipo textuais com que está trabalhando, tanto na cultura de partida quanto na de chegada. Nesse sentido, é fundamental que tenha também uma disposição (e habilidade) para pesquisar – o que vai lhe possibilitar reconhecer e preencher as possíveis lacunas de seu conhecimento especializado (que o dicionário não puder preencher) para, então, produzir um texto preciso e natural. Ou seja, um bom tradutor deve, acima de tudo, ser um bom produtor de textos – daí a importância do estilo e da criatividade na tradução especializada.

Mas, que implicações tem isso para um modelo de dicionário técnico? Seria o caso de repensar a Terminologia como um todo para ampliar seu objeto de estudo para além dos termos, fraseologias terminológicas, conceitos e definições? Ou seria o caso de empreender novas pesquisas no âmbito dos Estudos da Tradução (ou mesmo da Terminologia) que refletissem sobre possíveis modelos de “dicionários terminológicos especiais”, a exemplo do que ocorre com a Lexicologia na língua geral, voltados para as necessidades de um público-alvo específico: o tradutor? Essa nova linha de pesquisa – “Lexicologia Especializada”, de acordo com autores como Humbley (1997) e Valente (2002), mas que eu proporia chamar de **Terminologia Tradutiva** – procuraria compreender melhor a interface profícua entre essas duas áreas, fazendo a ponte necessária entre terminólogos e tradutores via Lingüística de Corpus (sobre a qual

falamos mais adiante). Dedicar-se-ia a estudar e testar novas metodologias de pesquisa para a identificação e descrição das unidades de tradução recorrentes em *corpora* de textos especializados – unidades essas que importam tanto à tradução, ainda que não necessariamente à terminologia (no sentido estrito da representação estruturada dos conceitos de uma dada área especializada).

Um exemplo concreto: dois dicionários bilíngües de Culinária disponíveis no Brasil atualmente (Klie, 2006; Moura, 2000) oferecem um equivalente para “pimenta-do-reino” em inglês – “black pepper” (aliás, até mesmo os dicionários de língua geral o fazem). Mas como dizer, em inglês: “pimenta-do-reino em grão(s)”, “pimenta-do-reino moída na hora”, “sal e pimenta-do-reino a gosto”, “uma (generosa) pitada de pimenta-do-reino”, “pimenta-do-reino branca”, “pimenta-do-reino verde”, “pimenta-do-reino moída grosseiramente”, “tempere com (a) pimenta-do-reino”, “salpique / polvilhe com (sal e) pimenta-do-reino”, “acerte o sal e a pimenta(-do-reino)”?

O mesmo se aplica a processos – numa tradução, o equivalente oferecido para “(to) slice” em ambos os dicionários é “fatiar” (e, por analogia, deduz o tradutor, “sliced” = “fatiado” e “slice” = “fatia”). Ambos desprezam o fato de que, em português, haver também o equivalente “(cortar / cortado) em rodela(s)” (e, analogamente, “rodela(s) de tomate / cebola / pimentão” etc.) – haveria uma diferença no **uso**? Ou seja, “fatiar / fatiado” ocorreria com certos tipos de alimentos e “(cortar / cortado) em rodela(s)” com outros?; algum deles seria mais recorrente que o outro?; um seria mais usado em uma parte da receita que em outra (no modo de fazer, no título ou na lista de ingredientes, por exemplo)?

Bem, se fossem apenas esses os padrões lingüísticos alvo da descrição proposta nesta pesquisa, não seria necessário buscar novas metodologias para dar conta desses fenômenos, uma vez que a maioria desses padrões já foi descrita em textos técnicos e vem sendo estudada por alguns pesquisadores em Terminologia – como é o caso, por exemplo, de Bevilacqua (1996 e 1999), que descreveu as Unidades Fraseológicas Especializadas (UFE); Estopá (1999), que trata das Unidades Terminológicas Poliléxicas (UTP) e Valente (2002), que trata de adjetivos e verbos do ponto de vista da Lexicologia Explanatória e Combinatória (Mel’čuk *et al.* 1995). Em todos esses estudos, no entanto, uma das premissas seguidas na identificação de uma unidade é a associação obrigatória a um termo. De acordo com Bevilacqua (2001: 107), no caso das UFEs, ao termo associa-se um “elemento eventivo” (verbal ou

procedente de verbo, p. ex.: *instaurar inquérito, cometer um delito, captação de energia*) e, no caso das UTPs, a combinatória tem um “caráter denominativo e valor referencial”, representando um nódulo de conhecimento na estruturação conceptual de um âmbito especializado, por exemplo: *energia eólica, energia nuclear*.

Supondo que a terminologia corresponda, de fato, a apenas 5-10% de um texto técnico (como sugere Newmark, 1988), em que tipo de material o tradutor deve se apoiar ao lidar com os 90% restantes? Nos dicionários de língua geral? Nos manuais de estilo, geralmente voltados para a língua geral? Seria essa porção majoritária do texto, de fato, condicionada pelas mesmas regras da língua geral – ou seja, o tradutor, em posse de um bom dicionário técnico deveria ser capaz de, usando seus bons conhecimentos de língua geral, produzir uma tradução numa língua de chegada que seja ao mesmo tempo precisa (para isso usaria o bom dicionário técnico) e natural (para isso usaria seus bons conhecimentos de língua, ou um dicionário de língua geral, e/ou um manual de estilo)?

Para mim, a resposta a essa pergunta é “não”. Minha hipótese é a de que o tradutor precisa conhecer não só as características terminológicas do gênero e do tipo textual com que está trabalhando nas línguas de partida e de chegada, mas também como a língua geral se configura em textos técnicos daquele gênero e tipologia em ambas as línguas. Ou seja, minha tese é de que um dicionário que seja útil ao tradutor deve apresentar informações não apenas relativas aos termos e conceitos de uma dada área especializada (o que tem sido feito, com maior ou menor êxito, usando-se as metodologias de trabalho da Terminologia), mas também informações relativas à padronização textual em um ou em diversos gêneros e tipos textuais dessa mesma área. Isto é, deve descrever as regularidades lingüísticas que tipificam esses textos, ainda que as mesmas compreendam fenômenos que, por não estarem diretamente relacionados com a terminologia da área *stricto sensu*, têm sido vistos apenas como manifestações da “língua geral” e, portanto, desconsiderados pelas obras terminográficas.

Por exemplo: em receitas culinárias usa-se “enquanto isso” (e não “nesse meio tempo”, “nesse ínterim”) para introduzir um procedimento que será executado enquanto se espera a conclusão de um outro, iniciado anteriormente. Mas nem “enquanto”, nem “isso”, e nem a combinação dos dois itens lexicais teriam lugar numa estrutura representativa do conhecimento especializado da área da Culinária. O mesmo poderia

ser dito de expressões como “um(a) a um(a)”, “não muito grossas”, “demais ingredientes”, “a gosto”, “quanto baste”, “com a ponta dos dedos”, “em local fresco e seco” ou “de um dia para o outro”.

Conforme demonstrou Estopá (1999), cada usuário das linguagens especializadas (p. ex.: documentador, tradutor, terminólogo, profissional) identifica como “termo” porções distintas e de tamanhos variados no texto técnico (Estopá, 1999: 177). O tradutor, em sua prática, não segmenta o texto em “termos” e “conceitos”, mas em unidades textuais, ou “unidades de tradução”². E é para essas unidades de tradução, que estão vinculadas a situações de comunicação específicas – uma para o texto de partida, outra para o texto de chegada –, que ele precisa de equivalentes. Minha hipótese é a de que essas unidades têm estruturas mais ou menos fixas e ocorrem regularmente em textos especializados de uma mesma área, gênero e/ou tipologia textual – fenômeno semelhante ao que ocorre na língua geral e que é chamado de **padronização** em Lingüística de Corpus (vide Cap. 3, item 3.3.3).

Ora, para identificar e compilar esses padrões típicos da área em dicionários técnicos é preciso que essa nova abordagem, voltada para o tradutor e para a produção textual na língua de chegada, empregue uma metodologia diferenciada da que vem sendo utilizada na compilação da maioria dos dicionários bi- ou multilíngües disponíveis atualmente nos pares de línguas que incluem o português brasileiro. Uma metodologia que possibilite a coleta e identificação de unidades que, independentemente de conterem um termo (ou representarem um conceito-chave da área) ou não, são recorrentes em textos de um determinado gênero ou tipo textual e, portanto, necessárias à tarefa tradutória.

Em seguida, é preciso que essas unidades sejam relacionadas a seu contexto lingüístico imediato e caracterizadas quanto à situação de comunicação em que se inserem (“só a parte branca”, por exemplo, é mais usado em que seção da receita culinária?; co-ocorre com que tipo de alimentos?; a “parte branca” de determinado alimento é usada na Culinária de ambas as culturas?³; é fácil de ser adquirida na outra

² Por ora, entendidas como um segmento de texto que o tradutor trata como uma unidade de sentido, que pode – ou não – coincidir com um termo (trato com mais vagar desse conceito no capítulo 1, quando falo sobre as Teorias da Tradução).

³ Haja vista, por exemplo, o emprego do talo da erva-cidreira no Brasil, introduzido com a culinária tailandesa (a propósito, referido também como *capim-santo* ou *capim-limão* quando empregado dessa forma): até bem pouco tempo, apenas as folhas dessa planta eram usadas por aqui, e somente no preparo de chás.

cultura, ou seria o caso de se oferecer um substituto?; e quais seriam seus possíveis substitutos?).

Aqui fica claro também que o especialista, peça-chave em todas as metodologias de trabalho em Terminologia, teria outro papel na elaboração de um dicionário desse tipo. Ele não mais seria o responsável pela aprovação ou não dos itens escolhidos para compor os verbetes do dicionário, ou pela validação ou correção das definições elaboradas pelo terminólogo, uma vez que termos e conceitos não são o foco desse tipo de obra. Auxiliaria, sim, na tarefa de estabelecer equivalentes para padrões lingüístico-terminológicos cuja equivalência interlingual seja parcial ou inexistente. Por exemplo: nos países de cultura inglesa usa-se, com freqüência, um termômetro para definir o ponto de cozimento de carnes; no Brasil, isto não é comum (embora seja possível, já que esse tipo de termômetro está se tornando mais popular) – que outras opções há para testar o ponto de cozimento de uma carne em receitas brasileiras?; e no caso da versão, a que temperatura corresponderia um ponto de cozimento descrito em português como “dourada por fora mas ainda rosada por dentro”?

O especialista vai contribuir, portanto, com seu conhecimento especializado para que seja possível, lingüisticamente, oferecer um equivalente, ou possíveis equivalentes, ou uma nota de rodapé explicativa que possa ser usada pelo tradutor-produtor-de-textos em seu texto de chegada. Nesse sentido, o especialista colaborador de um projeto dessa natureza teria, idealmente, um bom conhecimento também da segunda língua e cultura do par escolhido, ou seria um profissional da área falante nativo da língua de chegada e com conhecimento da língua de partida. Considerando-se que cada língua de um dicionário bilíngüe bidirecional assume a condição de língua de partida ou de chegada dependendo da direção em pauta, seria importante, se possível, contar com dois especialistas com essas características.

Proponho que o foco de um dicionário voltado para o tradutor não mais seja o “conteúdo especializado” de uma dada área, mas a forma como os textos que veiculam esse conteúdo se organizam lingüisticamente nas culturas em estudo. Assim, ainda que reconheça a importância de se dominar o significado e o equivalente de termos para a tradução acurada de receitas e de outros textos culinários, me concentrarei em observar o entorno dessas palavras: o como, quando e quanto elas ocorrem em textos reais da área, com que outras palavras, sintagmas e características textuais e

discursivas se relacionam, se sua frequência é ou não significativa nesses textos a ponto de merecer registro. A problemática da definição e das relações conceituais será deixada a cargo da Terminologia. O enfoque da pesquisa será em estudar os padrões léxico-gramaticais da linguagem especializada em textos autênticos da área, e em estabelecer (possíveis) relações de equivalência interlingual entre eles.

O critério de seleção das unidades que comporão o dicionário é, a meu ver, o grande diferencial dessa proposta, pois se pauta não mais pela pertinência ou não a um esquema estruturado de representação conceptual de um dado âmbito especializado, mas simplesmente por sua ocorrência significativa no *corpus* e, portanto, por seu uso na produção textual especializada da língua de chegada. Daí decorre a importância de compilar os *corpora* de estudo a partir de textos reais e de uso mais corrente na área, ao contrário da prática que muitas vezes privilegia a seleção de textos pela quantidade de termos, definições e/ou contextos explicativos que contêm.

Para identificar e compilar unidades que tipificam o uso da língua nos textos técnicos – que vou chamar de **Unidades de Tradução Especializadas** (doravante UTEs) –, a **Lingüística de Corpus** (doravante L.C.) é a área de estudos que julgo fornecer os subsídios teóricos e metodológicos mais adequados, pois “trabalha dentro de um quadro conceitual formado por uma abordagem empirista e uma visão da linguagem como sistema probabilístico” (Berber Sardinha, 2004: 30). Essa visão da linguagem, que se opõe à visão apriorística do racionalismo, privilegia a identificação de padrões léxico-gramaticais recorrentes por meio da observação empírica de **textos autênticos** organizados sob a forma de um *corpus* eletrônico. Esses padrões são então relacionados aos contextos comunicativos e tipos textuais em que ocorrem (veja, p. ex.: Sinclair, 1991; Barlow e Kemmer, 1994; Hunston e Francis, 1998, 1999 e 2000).

Descrita dessa maneira, a Lingüística de Corpus parece assemelhar-se muito, do ponto de vista metodológico, à Terminologia: ambas privilegiam o uso de *corpora* eletrônicos constituídos de textos autênticos e representativos para o levantamento de dados – aliás, a L.C. vem sendo usada, com sucesso, como metodologia auxiliar na compilação de obras terminográficas por vários pesquisadores (Dancette & Réthoré, 2000; Pérez Hernandez, 2002; Gómez Gonzáles-Jover, 2005; Fromm, 2007). No entanto, pretendo mostrar na presente pesquisa que a **visão de língua** que está por trás de cada uma delas, além de diferenciá-las teoricamente, permite apontar a L.C.

como a abordagem mais adequada à descrição do texto técnico do ponto de vista da tradução⁴.

Assim, este trabalho procurará demonstrar que o uso da Lingüística de Corpus, não apenas como metodologia, mas como **abordagem teórica** na exploração de *corpora* especializados pode contribuir para a elaboração de obras de referência mais úteis e confiáveis para o tradutor pois, em vez de considerar o texto técnico como um repositório de termos e definições, procura dar conta de quaisquer padrões associativos entre palavras que tenham uma alta probabilidade de ocorrência em textos representativos da área – fato que, na minha opinião, justifica plenamente a inclusão desses padrões num dicionário técnico voltado para o tradutor-produtor-de-textos.

Cumpri uma série de etapas intermediárias a fim de alcançar o objetivo principal de propor o modelo de dicionário *online* bidirecional português-inglês voltado para o tradutor da área: i) a compilação de um *corpus* comparável (originais nas duas línguas) de Culinária português / inglês, com 1,5 milhão de palavras em cada língua; ii) a descrição de padrões léxico-gramaticais típicos da área; iii) o desenvolvimento de metodologias de identificação semi-automatizada de UTEs em textos da área técnica da Culinária e iv) a elaboração de um modelo de verbete para o dicionário, com base na abordagem proposta – etapas que passo a relatar nos capítulos que se seguem.

O primeiro capítulo situa a Tradução Técnica no panorama teórico dos Estudos da Tradução, definindo alguns conceitos básicos da área como: unidade de tradução, equivalência, estratégias de tradução e naturalidade. Na seqüência, trato da Terminologia da perspectiva da Tradução Técnica, levando em conta algumas implicações práticas dessa interface para o dia-a-dia do tradutor especializado.

O segundo capítulo começa com um exame das especificidades do trabalho do tradutor técnico da área da Culinária no Brasil ressaltando seus campos de atuação, sua relação com as editoras, as fontes de consulta que tem à sua disposição e as lacunas dessas fontes. Em seguida, procuro caracterizar a “receita culinária”, alvo deste estudo, como gênero e tipologia textual, ressaltando suas principais características quanto ao seu meio de divulgação, sua organização e aparência, seus componentes estruturais e lingüísticos, grau de especialização e estilo. Por fim, abordo alguns problemas específicos da tradução no par de línguas português brasileiro-inglês.

⁴ essa diferença será discutida com mais vagar no Capítulo III.

A L.C. é apresentada, no terceiro capítulo, como abordagem teórica e metodológica a ser seguida na realização deste trabalho. Após um breve histórico, defino os principais conceitos da área pertinentes a esta pesquisa. Em seguida, apresento algumas metodologias da L.C. que vêm sendo usadas especificamente em pesquisas teóricas e práticas sobre a tradução técnica. Por fim, falo sobre as etapas de compilação do *corpus* alvo deste estudo.

No quarto capítulo, descrevo os métodos de etiquetagem e exploração do *corpus*. Dois estudos são apresentados: i) um levantamento manual da padronização de uma palavra-chave da Culinária em parte do *corpus* de inglês e o cotejamento dos resultados com as informações disponíveis em dicionários da área publicados no Brasil; ii) um estudo das entradas do glossário proveniente de minha memória de tradução Culinária sob várias perspectivas e a posterior extração semi-automática de padrões morfossintáticos semelhantes no *corpus* de português. Por fim, são discutidos os resultados obtidos na extração manual e semi-automática dos padrões.

No quinto capítulo proponho um dicionário *online* de Culinária voltado para o tradutor técnico, que servirá, futuramente, para abrigar as UTEs extraídas do *corpus*. Além de descrever as características de sua macro- e microestrutura, são destacadas as funcionalidades de sua interface de consulta *online* e a ficha a ser usada para a recolha dos dados é apresentada. Em seguida, são dados exemplos de verbetes nas duas direções tradutórias.

Por fim, discuto a validade dos métodos e dos resultados obtidos e suas implicações para a teoria e a prática da tradução técnica. À guisa de conclusão, aponto alguns desdobramentos possíveis de meu trabalho e sugiro pesquisas futuras que viriam auxiliar na descrição e compreensão do complexo fenômeno da interface Tradução Técnica - Terminologia à luz da Linguística de Corpus.

Espero, neste percurso, atendendo ao apelo das pesquisadoras citadas na epígrafe desta Introdução, repensar a compilação de dicionários especializados na era da informação digitalizada do ponto de vista das necessidades do tradutor técnico. Além disso, espero oferecer uma contribuição prática e concreta para os profissionais e aprendizes que trabalham com a tradução técnica no Brasil e, especialmente, com a tradução culinária – uma área que está se tornando cada vez mais especializada e com uma demanda crescente de traduções no mercado editorial nacional.

This document was created with Win2PDF available at <http://www.daneprairie.com>.
The unregistered version of Win2PDF is for evaluation or non-commercial use only.

Capítulo I

Tradução Técnica: teoria(s) e algumas questões práticas da interface com a Terminologia

1.1 Tradução técnica: prática e reflexão sobre a prática

A tradução técnica, em seu sentido lato de prática da comunicação interlingual, é tão antiga quanto a história do homem. A reflexão sobre essa prática, no entanto, teve início bem mais recentemente.

Alguns autores, como Bassnett-McGuire (1980), apontam como marco inicial das reflexões sobre a Tradução obras de pensadores romanos, especialmente Cícero (106 a 43 a.C.) e Horácio (65 a 8 a.C.), em que os autores tecem comentários acerca de suas escolhas ao verterem textos, majoritariamente literários e filosóficos, do grego para o latim. Outros teóricos, como Hung e Pollard (1998) e Krishnamurthy (1998), atentam para o fato de que, para além do panorama europeu, a tradução começou a

institucionalizar-se muito antes – desde o século IX a.C. na China, e desde o século IV a.C. no subcontinente indiano, respectivamente – embora haja poucos estudos publicados sobre o tema até o momento (Malmkjaer, 2005: 2).

Um segundo marco importante para o desenvolvimento do panorama teórico da disciplina no ocidente é a tradução da Bíblia – a primeira delas feita por São Jerônimo (340 – 419-20 d.C.), em cuja data de morte, 30 de setembro, celebra-se o Dia Internacional do Tradutor. Os debates teóricos suscitados pela prática da tradução bíblica, permeada de interesses religiosos e políticos, trouxeram à baila definitivamente a problemática já aludida no texto de Cícero, *De Oratore* (I: 35, *apud* Lefevere, 1990: 23-24), e que, de alguma forma, ainda está no cerne das teorias da modernidade: a fidelidade ao sentido ou à forma na tradução.

De um lado havia os defensores da tradução palavra-por-palavra, como o próprio São Jerônimo, como forma de preservar o mistério sagrado das escrituras; de outro, aqueles que defendiam a interpretação do texto sagrado pelo tradutor, que deveria então recodificar o sentido da mensagem divina de forma que pudesse ser compreendida pelo leitor, ainda que isso significasse uma transgressão da gramática do texto original, como é o caso de Wycliffite (1320-1384) e Lutero (1483-1546) – este, considerado o primeiro defensor da tradução voltada para o leitor (Robinson, 1997: 84).

Charles Batteaux (1715-1780) foi o primeiro autor a aproximar-se mais de uma formalização teórica propriamente sobre a tradução, procurando justificar sua detalhada lista de princípios com argumentos como, por exemplo, uma suposta universalidade dos padrões de pensamento. Em seu *Principes de littérature* (1747-8), Batteaux defendia, entre outras coisas, que as marcas de parágrafo do original, assim como as conjunções, devem ser sempre mantidas, que os advérbios devem ficar próximos dos verbos que modificam, respeitando a construção sintática do original, e que o tom, os tropos e os provérbios devem ser sempre reproduzidos fielmente (Malmkjaer, 2005: 7).

A partir do final do século XVII, foram surgindo reflexões teóricas cada vez mais elaboradas acerca da natureza da tradução, bem como sugestões de princípios norteadores para a sua boa prática, ainda que a tradução literária e bíblica fossem sempre o objeto de estudo privilegiado dessas observações.

Com os avanços teóricos ocorridos no campo da Lingüística a partir do começo do século XX, autores como Mounin (1963) e Jakobson (1969 [1960]) começaram a

discutir a tradução do ponto de vista dos estudos lingüísticos, especialmente da perspectiva estruturalista. Os modelos de descrição dos componentes da comunicação verbal de Jakobson – *emissor, receptor, canal, código, referente e mensagem* – e respectivas funções da linguagem (*emotiva, conativa, fática, metalingüística, referencial e poética*), bem como sua divisão tripartite dos tipos de “tradução” (em seu sentido lato) – *intralingual* (ou paráfrase), *interlingual* (tradução propriamente) e *intersemiótica* (ou transmutação, de um sistema de significação a outro) –, são citados e usados ainda hoje por muitos modelos teóricos da tradução.

A exemplo do que ocorreu em vários campos do saber, o Estruturalismo – e, posteriormente, outras abordagens teóricas, como o Funcionalismo, o Gerativismo, o Cognitivismo e a Sociolingüística – influenciaram os estudos lingüísticos como um todo, e os efeitos disso se fizeram sentir também nos estudos da tradução. Uma visão mais empírica do fenômeno tradutório contribuiu para o estabelecimento definitivo dos Estudos da Tradução como disciplina, na medida em que possibilitou a adoção de metodologias de pesquisa descritivas, baseadas na observação – especialmente do texto traduzido.

A origem do termo Estudos da Tradução (*Translation Studies*, em inglês) para designar uma área de estudos independente, preocupada com “o conjunto de problemas associados ao ato de traduzir e às traduções”⁴ é creditada, por vários teóricos, a James Stratton Holmes (1924-1986) – poeta, tradutor e acadêmico americano que viveu boa parte de sua vida na Holanda. Seu mapeamento da disciplina (vide Fig. 1), apresentado pela primeira vez no *Third International Congress of Applied Linguistics* de Copenhague, em 1972, influenciou o trabalho de muitos estudiosos da tradução – como Gideon Toury, André Lefevere, Anton Popovic, Itamar Even-Zohar e Raymond van den Broek –, embora tenha sido publicado em forma de artigo apenas em 1988 (Malmkjaer, 2005: 17).

⁴ “the complex of problems clustered around the phenomenon of translating and translations” (Holmes, 1988 [1972]: 9)

Figura 1: Estudos da Tradução para Holmes (baseado em Baker 1998: 278)

Vários estudiosos (Baker, 1998: 279, Ulrych, 1997: 421, entre outros) afirmam que Holmes foi responsável não apenas pela organização primeira da disciplina (visível na abrangência e detalhamento das subdivisões por ele sugeridas), mas também, e por outro lado, pela ampliação de seu escopo. Holmes enfatizou a natureza empírica e teórico-prática dos Estudos da Tradução, destacando sempre a relação dialética, de complementaridade e interdependência mútua que há entre as subáreas desse campo do saber. Além disso, os estudos da tradução passaram a contemplar de forma mais sistemática, entre outras coisas, a tradução técnica, os materiais de apoio do tradutor, o ensino e a crítica da tradução – muito pouco estudados e relegados a segundo plano até então.

A partir de Holmes e outros autores contemporâneos seus, os Estudos da Tradução⁵ passaram a ter uma autonomia maior em relação a outras disciplinas – como a Lingüística, os Estudos Literários e os Estudos Culturais – e a Tradução Técnica, enfim, passou a ter seu lugar enquanto objeto de estudo teórico, descritivo e aplicado.

At one time, the term ‘translation studies’ implied more emphasis on literary translation and less on other forms of translation, including interpreting, as well as lack of interest in practical issues such as pedagogy, but this is no longer the case. ‘Translation studies’ is now understood to refer to the academic discipline concerned with the study of translation at large, including literary and non-literary translation, various forms of oral interpreting, as well as dubbing and subtitling⁶. (Baker, 1998: 277)

O texto original foi, aos poucos – e em maior ou menor grau, dependendo da linha teórica –, deixando de ser o centro das atenções dos Estudos da Tradução, especialmente no que diz respeito à tradução técnica. Isso lançou nova luz ao antigo debate entre *fidelidade à forma vs. fidelidade ao sentido*, na medida em que os fatores envolvidos nas condições de produção e recepção do texto traduzido passaram a ser equacionados, trazendo à baila outras noções, tais como a de *aceitabilidade* (Toury, 1995: 57).

(...) a number of theorists began to advocate a shift of attention away from an exclusive occupation with the discovering ways of representing a Source Text faithfully in a translation towards viewing a translation as a factual phenomenon displaying a number of relationships, which might be called relationships of translation equivalence, with the Source Text⁷. (Malmkjaer, 2005: 3)

⁵ Outras terminologias foram propostas para nomear essa área de estudo, tais como “science of translation”, “translatology” e “traductologie” (Baker, 1998).

⁶ Houve um tempo em que o termo “estudos da tradução” implicava mais ênfase na tradução literária e menos em outros tipos de tradução, inclusive interpretação, bem como uma falta de interesse por aspectos práticos como o ensino da tradução. Mas isso mudou. ‘Estudos da Tradução’ refere-se, atualmente, à disciplina acadêmica preocupada em estudar a tradução em seu sentido lato, incluindo traduções literárias e não-literárias, várias formas de interpretação oral, além da dublagem e da legendagem.

⁷ (...) vários teóricos começaram a defender um deslocamento do foco de atenção da ocupação exclusiva com as formas de representar fielmente o Texto Fonte numa tradução para uma visão da tradução enquanto fenômeno factual que apresenta uma série de relações, que poderiam ser chamadas relações de equivalência tradutiva com o Texto Fonte.

The focus in translation investigation is shifting from the abstract to the specific, from deep underlying hypothetical forms to the surface of texts with all their gaps, errors, ambiguities, multiple referents, and “foreign” disorder. These are being analyzed – and not by standards of equivalent / inequivalent, right/wrong, good/bad, and correct/incorrect⁸. (Gentzler, 1993: 4)

Várias abordagens teóricas se estabeleceram no âmbito dos Estudos da Tradução ao longo desses últimos 40 anos – lingüísticas, comunicativas, funcionalistas, culturais, cognitivistas, para mencionar apenas algumas. Ressaltarei, a seguir, as contribuições que algumas delas deram à reflexão sobre uma prática específica: a tradução técnica.

1.2 A Tradução técnica no âmbito das teorias da tradução

Já afirmava Nord (1997: 29) que nenhuma teoria sobre a tradução – livre, literal, formal, dinâmica – explica completamente a tradução técnica. E, uma vez que não se tem notícia, até o momento, de uma Teoria da Tradução Técnica, é preciso pinçar contribuições a essa faceta da atividade tradutória em meio a pesquisas e reflexões realizadas sobre o tema em seu sentido mais lato. Mas isso não é tarefa fácil, já que a tradução técnica, como pontua Jody Byrne (2006), nunca esteve no centro das reflexões sobre a tradução: sempre foi a “prima pobre”, o “patinho feio”, “regarded as little more than an exercise in specialised terminology and subject knowledge”⁹ (p. 1) – ainda que seja responsável por uma enorme fatia do mercado de tradução há algumas décadas.

(...) technical translation, like other specialised types of translation, does not fit neatly into any one theory or approach and (...) there is, as yet, no adequate explanation of technical translation as an activity¹⁰. (Byrne, 2006: 2)

⁸ O foco da pesquisa sobre tradução está mudando do abstrato para o específico, das hipotéticas formas profundas subjacentes para a superfície textual com todas as suas lacunas, erros, ambigüidades, múltiplos referentes e estrangeirismos. Textos estão sendo analisados – e não segundo padrões de equivalente/não-equivalente, certo/errado, bom/ruim e correto/incorreto.

⁹ considerada meramente como um exercício em terminologia e conhecimento especializado

¹⁰ (...) a tradução técnica, assim como outros tipos de tradução especializada, não se encaixa perfeitamente em nenhuma teoria ou abordagem, e não foi formulada, até o momento, explicação adequada da tradução técnica enquanto atividade.

Byrne (2006: 2-7) elenca várias crenças errôneas (“*misconceptions*”) a respeito da tradução técnica que, para ele, impedem o reconhecimento da verdadeira importância e complexidade da natureza dessa atividade. Algumas dessas crenças – ou, nas palavras do autor, “mitos” – são:

- na tradução técnica, só o que importa é a terminologia;
- o estilo não tem importância na tradução técnica;
- a tradução técnica não envolve criatividade – é simplesmente um processo de transferência por reprodução;
- para ser um bom tradutor técnico, é preciso ser um experto de uma área altamente especializada;
- a tradução técnica se reduz à transmissão de informação especializada.

Outra estudiosa da tradução técnica, Karen Korning Zethsen (1999, 1997), chegou a uma lista semelhante de itens – que chamou de “dogmas” – após coletar comentários sobre a natureza da tradução técnica citados em trabalhos de autores como Newmark (1988: 10, 151), Myers (1989: 3), Bell (1991: 203), Nord (1991: 178, 182), Baakes (1994: 5), Wilss (1996: 22-23) e Laurén (1998: 463):

To sum up technical translation seems to be subject to the following dogmas:

1. The purpose of the translation is to transmit factual information.

2. The greatest problem is terminology.

3. No particular translation strategy is needed as long as the translator is familiar with the relevant terminology and the typical syntax of a technical text.

(...) the stereotypical view of the technical writer, the focus on terminology and the convenient polarisation literature / science-technology have all contributed to the rather narrow perception of technical texts which has prevailed for many years. It goes without saying that the dominant view of technical translation has been equally narrow.¹¹ (Korning Zethsen, 1999: 69, grifo da autora)

¹¹ Resumindo, a tradução técnica parece estar sujeita aos seguintes dogmas:

1. O propósito da tradução é transmitir informação factual

2. O maior problema é a terminologia

3. Nenhuma estratégia de tradução em particular é necessária se o tradutor estiver familiarizado com a terminologia relevante e a sintaxe típica do texto técnico

(...) a visão estereotipada do redator técnico, o foco na terminologia e a conveniente polarização literatura / ciência-tecnologia contribuíram para uma percepção deveras limitada do texto técnico que prevalece há anos. Desnecessário dizer que a visão dominante sobre a tradução técnica é igualmente limitada.

Como é possível depreender nas observações de Byrne e Korning Zethsen, corroboradas por outros autores – como Azenha Jr. (1999) – o texto técnico, do ponto de vista da tradução, parece ter sido sempre reduzido a seu conteúdo especializado, terminológico. Os autores concordam, por um lado, que isso prejudicou bastante a compreensão da tradução técnica e, conseqüentemente, sua práxis e seu estudo. Também são uníssonos em afirmar que somente uma visão mais abrangente do texto técnico será capaz de dar conta das múltiplas variáveis envolvidas na tradução desse tipo textual.

(...) creio que uma reflexão mais sistemática sobre a natureza e as características do texto técnico pode contribuir, de um lado, para lançar nova luz sobre as relações envolvidas nesse tipo de tradução e, de outro, para redefinir o status de que desfruta esse tipo de tradução no interior das discussões acadêmicas. (Azenha Jr., 1999: 9)

Nos subitens que se seguem, serão ressaltados alguns aspectos teóricos do trabalho desses e de outros autores que consideram o texto técnico não como mero veículo de informação especializada, mas como fenômeno socio-cultural, pragmático e discursivo.

1.2.1 Texto técnico e cultura

João Azenha Jr., em seu livro de 1999¹², argumenta em favor de uma visão de tradução técnica enquanto **fenômeno culturalmente condicionado**. Para ele, o texto técnico é “estrutura multidimensional ancorada historicamente e composta por diferentes planos inter-relacionados, todos eles portadores de sentido e, portanto, de relevância para o tradutor” (Azenha Jr., 1999: 11-12).

Sua obra divide-se em três capítulos. O primeiro apresenta os pressupostos teóricos – discute as relações entre linguagem e cultura e apresenta dois modelos de tipologia textual e sua relação com teorias sobre o texto técnico. O segundo analisa, de acordo com aporte teórico apresentado, as estratégias empregadas na tradução dos seguintes textos “instrutivos” alemão-português: i) folhetos informativos sobre fitas para gravador; ii) folhetos informativos e propostas de adesão a cartão de crédito; iii)

¹² Obra sobre tradução técnica que se tornou referência sobre o tema no Brasil, é uma reescritura de sua tese de doutorado, defendida junto ao Departamento de Letras Modernas da USP, área de Alemão, em 1994.

manuals informativos para a aplicação de fungicidas; iv) bulas de remédio; v) instruções para preparação e aplicação de tintura para cabelos e vi) manuals de proprietário de automóvel. O terceiro e último capítulo dedica-se à comprovação das hipóteses iniciais e aponta para desdobramentos futuros da pesquisa.

O autor faz um apanhado das teorias que dominaram os estudos tradutológicos dos anos 1970 a 1990 – sistematização que empresta de Frank-G. Königs, para quem “(...) a tradução não pode ser vista como uma simples substituição de signos, mas como uma transferência e adaptação cultural de conteúdos extralingüísticos” (Königs, 1992: 31 *apud* Azenha Jr., 1999: 30). Então, situa o surgimento do que, para Azenha Jr., foi o início de uma visão mais abrangente e cultural do fenômeno tradutório:

(...) foi na década de 80 que ganharam um impulso decisivo os estudos de tradução emoldurados por essa dimensão cultural (...) a partir da qual a tradução é vista como ‘um tipo especial de transferência de valores culturais’ (Reiss e Vermeer, 1984). (Azenha Jr., 1999: 9-10)

Cita o conceito de “cultura” formulado por vários estudiosos, entre eles o etnologista Ward H. Goodenough e os lingüistas Halliday e Hasan, que vêem a cultura não propriamente como um fenômeno material – pessoas, coisas, emoções ou comportamentos específicos – mas, antes, uma **forma típica de organizar** esses fenômenos, ou um pacote (“**bundle**”) de coisas cuja reunião é típica. Da mesma maneira, os “tipos textuais” caracterizariam-se pela forma típica com que certas peculiaridades lingüísticas e extratextuais se organizam.

Vários estudiosos da tradução têm se dedicado a descrever as peculiaridades das “tipologias textuais” da perspectiva da tradução. Azenha Jr. discute o conceito apresentando dois modelos teóricos: o de Reiss (1976) e o de Snell-Hornby (1988). O modelo de Reiss foca primordialmente a função comunicativa do texto de chegada e, segundo o autor, apresenta falhas. Para ele, é difícil definir a hierarquia das funções propostas pela autora (“informativa”, “expressiva” e “operativa”) em certos textos. Essa opinião está em consonância com a de outros autores que estudaram a natureza do texto técnico e de sua tradução, como Korning Zethsen, que comenta:

The translation-oriented text typologies have no doubt done much harm in relation to technical texts. The fact that technical texts are automatically classified as being informative has

greatly influenced the way in which technical texts have generally been perceived. Some might claim that the texts are only classified according to their dominant function, but one could question whether such a classification is possible. How does one assess which function is most important in a text – is it necessarily the one that takes up most space?¹³ (Korning Zethsen, 1999: 68).

Além disso, os métodos empregados por Reiss para descrever o modelo de tradução recomendado para cada tipo textual (“identificante”, “prosaico” e “adaptante”) são vagos, de acordo com Azenha. O autor critica também a proposição da autora de um ideal de “invariância semântica” supostamente almejado pelas traduções – como se fosse possível haver uma correspondência de sentido absoluta entre os textos de partida e chegada.

O segundo modelo apresentado por Azenha, de Snell-Hornby, é elogiado pela “contribuição que representa para uma redefinição de critérios a partir dos quais se podem nortear os estudos tradutológicos e, em última análise, a visão de tradução” e, especialmente, porque a autora considera a interdisciplinaridade e a intertextualidade como elementos indissociáveis do fazer tradutório (Azenha Jr., 1999: 60).

Snell-Hornby afirma que duas correntes têm dominado as reflexões sobre a tradução: a tradição de 2.000 anos centrada em obras literárias e bíblicas, e uma voltada para a tradução técnica, que considera o texto especializado como mero veículo informativo para a transmissão de conhecimentos técnico-científicos. A proposta da autora é contribuir para diminuir o abismo entre essas duas correntes repensando a forma de “conceituar” e “categorizar”.

Segundo ela, toda ciência precisa de categorias, e categorias precisam de conceitos, e os dois conceitos fundamentais de categorização que detectou nos estudos lingüísticos são a “dicotomia”, ou polarização rígida, e a “tipologia”, ou sistema de compartimentos estanques (*apud* Azenha Jr., 1999: 53). Para ela, uma teoria mais adequada para os estudos da tradução seria derivada da teoria da categorização natural, elaborada por Rosch (1973), que trabalha o conceito de **núcleo com extremidades esmaecidas** – já tratada por Wittgenstein em 1953, em seu

¹³ *As tipologias textuais baseadas na tradução, sem dúvida, foram muito prejudiciais ao texto técnico. O fato de textos técnicos serem automaticamente classificados como informativos influenciou sobremaneira o modo como esses textos vem sendo compreendidos de maneira geral. Alguns podem afirmar que os textos são classificados de acordo apenas com sua função dominante, mas pode-se questionar se tal classificação é mesmo factível. Como é possível determinar qual é a função mais importante num texto – seria, necessariamente, aquela que ocupa mais espaço?*

Investigações filosóficas (Wittgenstein, 1987 [1953]). Na teoria de Snell-Hornby, esse conceito recebe o nome de “princípio holístico do Gestalt” – ele permitiria uma abordagem integrada da tradução ao lançar mão de:

(...) pelo menos dois conceitos da Psicologia gestaltiana: (1) o todo é mais do que a mera soma de suas partes; (2) uma análise das partes não fornece uma compreensão do todo – conceitos ignorados por filólogos e lingüistas que estudaram a tradução, segundo a autora. (Azenha Jr., 1999: 52).

Assim como foi apontado com relação ao conceito de cultura discutido por Azenha Jr., essa noção de que **o todo é mais do que a mera soma das partes** está no cerne deste trabalho: qualquer tradutor com um mínimo de prática saberá que, para traduzir um texto técnico não basta ter em mãos um dicionário técnico, especialmente porque a teoria terminológica que está por trás desses dicionários tem se concentrado, historicamente, em certas “partes” que compõem o texto técnico (i. e., substantivos e, em menor escala, verbos e adjetivos), mas muito pouco tem se dedicado ao “todo” da comunicação especializada, ou à **forma típica** como essas “partes” **se organizam** no texto especializado, seja no nível micro- seja no nível macro-textual.

Azenha critica o trabalho de Snell-Hornby no que diz respeito à aplicabilidade de suas reflexões – para ele, o modelo não oferece metodologias muito claras para a prática tradutória em si. Para mim, no entanto, essas reflexões estão em consonância com a visão de texto técnico, de léxico especializado e de tradução que busco defender neste trabalho.

Para falar de tipologia do ponto de vista dos estudiosos das linguagens técnicas (LSP – Languages for Special Purposes), Azenha Jr. apresenta as teorias de Hoffmann (1985) e Wilss (1997), para quem as linguagens técnicas podem ser diferenciadas a partir de sua análise em dois planos – “horizontal” (em comparação com outras linguagens técnicas) e “vertical” (variações de registro). A respeito das categorias propostas pelos autores, Azenha comenta: “Wilss acredita ser o inventário lexical e, dentro dele os substantivos, o lugar de excelência em que as linguagens técnicas diferem da linguagem comum”. Essa afirmação reforça, mais uma vez, a visão “lexicocêntrica” da tradução técnica – criticada por Azenha e outros autores:

Technical terminology has been a favorite research area to LSP researchers and much important research has been carried out

within this field, but unfortunately this focus on terminology has also (according to Hoffmann 1991) lead to a state of lexicocentrism.¹⁴ (Korning Zethsen, 1999: 73).

Da perspectiva bilíngüe, o léxico especializado é apenas um dos fatores envolvidos na noção de “**equivalência interligual**”. Os dicionários técnicos, em geral meras listas com *termo LP* → *termo LCh*, têm evitado tratar dos casos de equivalência parcial e não-equivalência entre as línguas. Por que esses dicionários não oferecem, por exemplo, uma equivalência *termo LP* → ∅, se o tradutor, em sua prática diária, pode (e algumas vezes deve) optar por essa estratégia em determinados contextos?

Por exemplo: ao traduzir para o português uma receita que ofereça a medida de um ingrediente em *fluid ounces* (fl oz) e mililitros (ml), o tradutor deve omitir *fluid ounces* – e não traduzir por “onças líquidas”, como fizeram alguns aprendizes de tradução inexperientes no estudo de caso que realizei no mestrado (Teixeira, 2004: 132). Assim, ainda que “onças líquidas” possa ser considerado “equivalente” *lato sensu* em português para *fluid ounces*, em receitas culinárias, *fluid ounces* deverá ser convertido para “mililitros” caso ocorra sozinho, e, no caso de ocorrer lado a lado com uma medida em mililitros e/ou colheres / xícaras, deve ser omitido. Essa discrepância entre a informação oferecida por dicionários bilíngües e as necessidades e peculiaridades da tarefa tradutória chama a atenção de autores como Hunston (2002: 125): “dictionaries have a bias in favour of finding a lexical equivalent or paraphrases – but the more sucessfully they do this, the less faithfully they reproduce actual translator behaviour”¹⁵.

Azenha Jr. discute o conceito equivalência interlingual com base na tipologia proposta por Schmitt (1989). Para o autor, o modelo de Schmitt – ainda que contribua para demonstrar que “os problemas de tradução não se restringem ao domínio de uma terminologia, mas ao conhecimento de uma realidade lingüística e conhecimento de mundo” –, limita-se ao léxico, deixando a sintaxe, a estruturação geral do texto e o emprego de elementos gráficos e tipográficos de lado (recursos icônicos). Azenha chama a atenção para o fato de que as ilustrações, em muitos textos traduzidos,

¹⁴ A terminologia especializada tem sido uma área de estudo preferencial dos pesquisadores de LSP (Language for Special Purposes – Linguagens para Fins Especiais), e investigações importantes têm sido conduzidas nessa área; mas, infelizmente, esse foco na terminologia também tem conduzido (de acordo com Hoffmann, 1991) a um estado de lexicocentrismo.

¹⁵ dicionários têm uma tendência a privilegiar o estabelecimento de equivalentes lexicais ou paráfrases – mas, quanto mais bem sucedidos eles são nessa tarefa, menos representam o comportamento usual dos tradutores.

podem estar em desacordo com a cultura de chegada, produzindo uma incongruência (como, por exemplo, a foto de um ingrediente ou equipamento que não existe no Brasil, no caso da tradução de uma receita culinária). Comenta: “em cada texto, é preciso estabelecer um elo entre os elementos constituintes de sentido em todos os níveis de organização textual”.

Ao refletir sobre os resultados de sua pesquisa no capítulo final do livro, Azenha Jr. conclui que “a tradução de textos técnicos está sujeita a variáveis que são diferentes em grau, mas não em essência, das que influenciam a tradução de qualquer tipo de texto” (Azenha Jr., 1999: 122) e, por isso mesmo, a tradução técnica não pode ser reduzida à problemática da equivalência terminológica. Isso corrobora diretamente minha hipótese de que um dicionário técnico adequado à produção e compreensão textuais envolvidas na tarefa tradutória deve dar conta não somente da terminologia da área, mas de diversas variáveis lingüísticas e extralingüísticas e de como estas influenciam e tipificam o texto técnico quanto ao seu gênero, tipo textual e contexto cultural de produção.

Os problemas de tratamento do léxico na tradução técnica que o autor identifica em seu estudo são oriundos, segundo ele, da falta de uma padronização terminológica, da deficiência das fontes de consulta (dicionários e glossários especializados), da escassez de bancos de dados e do descompasso tecnológico entre as nações e, por extensão, entre as línguas (o que vai repetir, como mostro mais adiante, no caso específico da Culinária).

Entre as contribuições mais importantes do trabalho de Azenha Jr. para a presente pesquisa estão o levantamento pormenorizado que fez da complexidade de fatores envolvidos na tradução de textos técnicos e o delineamento do importante papel desempenhado pelo tradutor no controle e harmonização dessas variáveis. A caracterização circunstanciada do texto técnico instrutivo e de suas condições de produção e recepção levadas a cabo pelo autor comprovam empiricamente que este não pode – e não deve – ser reduzido a seu conteúdo terminológico.

Pouca validade teria para a qualidade da tradução técnica uma acuidade terminológico-conceitual que beirasse os 100%, se ela estivesse inserida em estruturas sintáticas complexas, organizadas em períodos longos, com uma pontuação deficiente, numa programação visual que em nada auxiliasse o leitor (...) (Azenha Jr., 1999: 133).

“(...) o conhecimento técnico não pode garantir sozinho a qualidade da tradução de um texto técnico, já que a produção e a tradução de um texto, técnico ou não, envolvem – isso sim – conhecimentos específicos sobre os **mecanismos de produção de sentido em linguagem**.” (op. cit.: 131, grifo meu)

Seus achados apontam para a necessidade que o tradutor tem, em sua prática diária, de ter acesso a materiais confiáveis que dêem conta não somente do conteúdo especializado, mas de outros tipos de informações lingüísticas e extra-lingüísticas culturalmente condicionadas que, até o momento, não foram sistematizadas e disponibilizadas nos materiais de consulta tradicionalmente destinados a esse profissional.

Azenha Jr. cita a *Culinária en passant* ao falar da importância do aspecto visual e dos recursos icônicos para o texto instrutivo (pp. 86-87), e é possível traçar alguns paralelos entre os tipos textuais trabalhados pelo autor e as receitas, tema deste estudo. Mas, pareceu-me bastante difícil encaixar a *Culinária* em quaisquer dos modelos, categorizações e exemplificações que emprestou de Hoffmann e Wilss para descrever as linguagens técnicas. Que tipo de linguagem especializada seria, afinal, a *Culinária*? Seria necessária uma Terminologia Cultural (ou, como propõe Barbosa, 2006, para o discurso literário, uma Etno-terminologia) para dar conta dos aspectos que diferenciam, por exemplo, a terminologia contida nas receitas da terminologia dos demais textos instrutivos?

É precisamente neste ponto que o este trabalho se distancia, também, do de Jody Byrne (2006), que comento no subitem a seguir.

1.2.2 Texto técnico e usabilidade

Jody Byrne, em seu livro, publicado em 2006, *Technical Translation: Usability Strategies for Translating Technical Documentation*¹⁶, assim como Azenha, tem no texto técnico, e mais precisamente no texto instrutivo "software user guide" (manual do usuário de software) seu tema de trabalho.

¹⁶ *Tradução técnica: estratégias de usabilidade para a tradução de documentação técnica*. O livro é também baseado em sua tese de doutorado: *Textual Cognectics and the Role of Iconic Linkage in Software User Guides*, defendida na Dublin City University em 2004, sob a orientação de Dorothy Kenny.

A obra divide-se em seis capítulos. O primeiro sumariza as principais teorias sobre a tradução no que diz respeito à tradução técnica e elege a abordagem comunicativa, centrada no texto traduzido – em especial a Teoria da Relevância (de Gutt) e a *Skopos Theory* (de Vermeer e Nord) – como as mais adequadas para pensar a natureza e a finalidade da tradução técnica. O Capítulo 2 trata da comunicação especializada e das características da documentação técnica, oferecendo uma descrição detalhada dos fatores envolvidos na redação de manuais de software, seus recursos textuais, icônicos e comunicativos.

Uma das primeiras distinções que o autor faz no capítulo 1 é entre a tradução técnica e a científica:

(...) we can say that scientific translation relates to pure science in all of its theoretical, esoteric and cerebral glory while technical translation relates to how scientific knowledge is actually put to practical use, dirty fingernails and all.¹⁷ (Byrne, 2006: 8).

Infelizmente, a Culinária, enquanto temática, não se encaixa perfeitamente em nenhuma dessas duas categorias, apesar de conter elementos usados por Byrne para caracterizar ambas: uma receita culinária é, ao mesmo tempo, “simple, unambiguous, concise, and, for want of better word, unremarkable” (p. 9) como o texto técnico, além de poder apresentar as informações “in an interesting, serious and impressive way” (p. 8), como o texto científico.

Seu livro interessa-nos, no entanto, pela visão que o autor tem, por um lado, da tradução técnica como um **ato comunicativo** e, por outro, do tradutor técnico como um tipo especial de **redator técnico**, em consonância com o que vimos em Azenha Jr., embora o aporte teórico empregado seja outro.

Para apresentar o conceito de “iconic linkage” (linkagem icônica?), definido por ele como *the use of isomorphic constructions to express what is, essentially, the same information*¹⁸, o autor passa todo o capítulo 3 e grande parte do capítulo 4 explicando uma série de teorias cognitivistas e behavioristas. Seu objetivo é descrever os

¹⁷ (...) podemos afirmar que a tradução científica relaciona-se com a ciência pura em toda a sua glória teórica, esotérica e cerebral, ao passo que a tradução técnica relaciona-se ao modo como o conhecimento científico é de fato colocado em prática, unhas encardidas e tal.

¹⁸ “o uso de construções isomórficas para expressar o que, essencialmente, é a mesma informação” – ou seja, quando a mesma informação for usada várias vezes deve ter exatamente a mesma formulação textual (Byrne, 2006: 164)

mecanismos de cognição do usuário de manuais de software, e como este assimila, interpreta e memoriza informações. Sua tese é a de que, por meio do uso da “linkagem icônica” é possível melhorar a “usabilidade” (*usability*¹⁹) desse tipo de texto técnico. Para comprovar sua hipótese, o autor apresenta um experimento prático no quinto capítulo em que avalia as melhoras obtidas com relação à usabilidade em parte de um manual do usuário de software após este ter sido reescrito usando-se as estratégias de “linkagem icônica” por ele levantadas.

Não fosse pelo título da obra e pelo capítulo inicial, poderíamos dizer que a obra de Byrne pouco tem a ver com a tradução técnica (ou talvez tenha, se vista de uma perspectiva mais cognitivista e processual da tradução). Os focos principais de seu trabalho são, sim, o *texto técnico* e a *redação técnica*, e isso fica patente nos demais capítulos teóricos e na parte experimental de sua pesquisa, que não tem qualquer relação com a tradução. Mas a visão de tradução técnica que escolheu privilegiar – e que, com algumas ressalvas, partilho no presente trabalho – justifica suas escolhas teóricas e metodológicas:

Our role [translators] as communicators means that we have a responsibility to our readers to ensure that the information is communicated effectively. This is particularly true in technical and instructional texts. (...) Looking at **technical texts as a communicative activity** rather than as a purely text-based, linguistic activity opens the door to a variety of new and interesting avenues of investigation²⁰. (Byrne, 2006: 253, grifo meu).

(...) if we want to translate technical texts effectively and achieve the highest standards of quality it is essential that we understand how these texts are produced in the first place (...) **translations must be produced to the same standards and expectations as “native” texts**²¹. (*op. cit.*: 257, grifo meu)

¹⁹ De acordo com a norma ISO 9241-11, *usability* é: “the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context.” [a medida que um produto pode ser usado por usuários específicos para alcançar objetivos específicos com efetividade, eficiência, satisfação e em determinado contexto.]

²⁰ *Nosso papel como comunicadores significa que temos uma responsabilidade para com nossos leitores de assegurar-lhes a transmissão eficiente da informação. Isso é particularmente importante em textos técnicos e instrucionais. (...) Se os textos técnicos forem vistos como uma atividade comunicativa em vez de mera atividade lingüística centrada no texto, novas portas se abrirão para uma série de novos e interessantes caminhos de pesquisa.*

²¹ (...) *se queremos traduzir textos técnicos de forma eficiente e alcançar os padrões mais altos de qualidade é essencial que entendamos como esses textos são produzidos em primeiro lugar (...) o texto traduzido deve ser produzido para atender os mesmos padrões e expectativas dos textos “nativos”.*

O autor estabelece essa identidade entre tradutor e redator técnico em vários momentos de seu texto, a exemplo do que fizeram autores como Amman e Vermeer (1990: 27) e Göpferich (1993), para os quais o tradutor é um redator técnico intercultural (“intercultural” / “cross-cultural technical writer”). Tal posicionamento chama a atenção também de Azenha, que cita Amman (1989: 46): “O tradutor / intérprete é um especialista em comunicação intercultural (...) e não um ‘mediador’ entre duas línguas”. E comenta, a esse respeito:

Essa posição de responsabilidade na administração de muitas variáveis redimensiona o papel do tradutor que (...) passa a participante ativo no processo de tradução, nele atuando como “técnico”, como “especialista” para a **produção de textos cujo objetivo é viabilizar a comunicação entre culturas**. Uma tal revalorização do papel do tradutor representa uma contribuição importante dessa linha de estudos tradutológicos para a profissionalização da atividade de tradução (...). (Azenha Jr., 1999: 38, grifo meu).

Outro aspecto que Byrne ressalta em sua obra é o papel que o leitor – ou “usuário”, de acordo com Holz-Mänttari (1984: 111) – do texto de instrução deve desempenhar nas decisões do tradutor técnico, já que sua proposta é, justamente, melhorar a “usabilidade” das traduções de manuais de software. A questão que se coloca (e que nome pergunto, já que a Culinária não pode ser considerada um “texto técnico” propriamente, de acordo com a definição de Byrne) é: o que esse “usuário” espera de um texto como esse, traduzido ou não? E, mais uma vez, o debate da fidelidade na tradução, que aludi nas primeiras páginas deste capítulo, vem à tona.

What the user wants from a translation has been subject to much debate, argument and hand-wringing among translation theorists. Some argue that the target language reader may be interested in gaining an insight into the culture of the source text culture or the source text language itself. Others would argue that the target reader wants to experience the text in the same way as the original audience did. In certain circumstances this may be true. However, **in the case of technical translation, all readers are concerned about is getting the information they need and being able to understand and use it effectively in order to do something else**, usually some task relating to their day to day work²². (Byrne, 2006: 15, grifo meu).

²² O que o usuário espera de uma tradução tem sido tema de muitos debates, discussões e apreensão entre os teóricos da tradução. Alguns argumentam que o leitor da língua de chegada pode estar

Essa afirmação do autor (que destaquei em negrito) não seria um “retorno” à idéia de que o texto técnico é meramente informativo? Até que ponto suas observações se aplicam à Culinária, mais especificamente às receitas? O que aproxima uma receita de um texto de instrução como os descritos por Azenha e Byrne, e o que a distingue desse tipo textual? Pretendo tratar disso mais adiante, no Capítulo II, quando for discutir a Culinária. Por ora, concordo com Byrne quando afirma que o tradutor que ignora ou desconhece as convenções da língua de chegada para aquele tipo textual pode destituir o texto, o autor e a própria informação veiculada pela tradução de sua credibilidade (Byrne, 2006: 4).

Para conhecer essas convenções da linguagem especializada é preciso observar textos semelhantes, que foram produzidos naturalmente por falantes nativos naquela língua e cultura, e estudar suas características. Desse ponto de vista, a “terminologia”, como já argumentei anteriormente, passa a ser apenas um dos muitos aspectos a serem levados em consideração: “(...) a translator should have less trouble locating appropriate specialised terms in the target language than with the non-specialised, general terms”²³ (Byrne, 2006: 4). Ora, se o texto técnico não pode ser reduzido à terminologia que veicula, me pergunto, mais uma vez, em que material de consulta o tradutor vai encontrar essas informações, para que possa traduzir todos os outros aspectos presentes no texto técnico? Se seu objetivo, como defende Byrne da perspectiva comunicativa (que nós, nesse sentido, endossamos), é produzir um texto com alto grau de “usabilidade”, o tradutor terá que mimetizar os textos produzidos pelos falantes nativos daquela língua e subespecialidade, sendo fiel ao leitor (e ao sentido do original) em sua tradução. E, para nós, é no dicionário técnico especializado que o tradutor deveria encontrar o máximo possível dessas informações.

interessado em conhecer um pouco da cultura do texto de partida ou da língua em si. Outros defendem que o leitor da tradução quer vivenciar o texto da mesma maneira que os leitores do texto original o fizeram. Em certos casos isso pode ser verdade. Entretanto, no caso da tradução técnica, a única coisa que de fato interessa aos leitores é obter a informação de que necessitam e poder entendê-la e usá-la efetivamente para realizar algo, em geral alguma tarefa relacionada à sua rotina de trabalho.

²³ (...) é provável que o tradutor tenha menos problemas para encontrar termos especializados apropriados na língua-alvo do que termos não-especializados, da língua geral.

1.2.3 Texto técnico e tipologia textual

Korning Zethsen (1999), a exemplo de outros autores citados até aqui, também argumenta que a tradução técnica vai muito além do conhecimento de uma terminologia. Para a autora, uma consequência indesejável desse “dogma”, segundo o qual a tradução técnica se reduz à transferência de um conteúdo especializado de uma língua a outra, é a idéia implícita de que textos técnicos têm como propósito único a transmissão de informação factual (sendo, portanto, puramente informativos).

Fortunately it has become increasingly clear to many scholars that literary texts do not hold a monopoly on expressivity and creativity. As early as in 1969 Newmark (1969: 80) writes: “...the common assumption that scientific and technical writing is concerned only with facts... is misplaced”²⁴. (Korning Zethsen, 1999: 72).

A autora argumenta que o conceito de “texto técnico” é falho, e tem sido aplicado a textos que não necessariamente caracterizam-se pela ausência de expressividade, pela imparcialidade, pela objetividade, etc., por muito tempo associadas ao texto técnico-científico. A autora faz uma proposição ousada: que as noções de gênero e tipo textual sejam postas de lado pelas teorias sobre a tradução técnica.

(...) in reality there is no such thing as a “technical text” – it is not a genre – but rather a number of texts making use of “technical language” and belonging to a great variety of genres²⁵. (Korning Zethsen, 1999: 74).

Para ela, um texto não pode ser classificado como “técnico” ou “não-técnico” só porque faz uso de uma “linguagem técnica”, pois isso seria o mesmo que afirmar que todos os textos que fazem uso dessa linguagem técnica pertencem a um mesmo gênero. O que dizer, por exemplo, de um texto literário que detalha, por quase quatro páginas, minúcias de um jogo de baseball (*Patrimony*, de Phillip Roth, 1991) – é um texto técnico? Ou um folheto informativo sobre uma casa de repouso para idosos – não

²⁴ Felizmente, tem se tornado cada vez mais claro para muitos acadêmicos que os textos literários não detém o monopólio sobre a expressividade e a criatividade. Já em 1969 Newmark (1969: 80) escreve: “... a suposição habitual de que os textos técnicos e científicos dizem respeito apenas a fatos... é errônea”

²⁵ (...) na realidade, o “texto técnico” não existe – não é um gênero, mas antes um conjunto de textos que fazem uso da “linguagem técnica” e pertencem a vários gêneros.

seria técnico, por usar a emotividade ou estratégias de convencimento? A própria receita culinária – não poderia ser considerada texto técnico por não veicular conhecimento científico-tecnológico (no sentido dado por Byrne)? Para a autora, uma classificação desse tipo apenas confunde o tradutor.

A proposta de Korning Zethsen é considerar o texto técnico, acima de tudo, como um **texto** – cada texto exige estratégias de tradução específicas, dependendo de suas características e das condições de produção e recepção tanto do original quanto do texto traduzido. Uma abordagem adequada para descrever o texto técnico dessa perspectiva seria, para a autora, a lingüística textual de Hoffmann (1991: 158).

The hunt for terminology has overshadowed the fact that **there is something between the terms which is equally important for a satisfactory translation** [grifo meu]. Instead of considering a technical text as a collection of technical terms it should be seen as a **text** [grifo da autora] – i.e. as an entity written to serve various purposes. Awareness of these purposes cannot be reached by focusing on terminology and by working from sentence to sentence (...) ²⁶. (Korning Zethsen, 1999: 71).

Para corroborar seu posicionamento, a autora cita um trabalho não publicado de Hatim²⁷ em que o autor sugere que o foco dos estudos da tradução voltados para o texto técnico, tradicionalmente centrado na terminologia e no grau de formalidade e/ou especialização desses textos, seja deslocado para:

“(...) those neglected aspects of textual practices (the discourse, the genre and **the actual texts** that typify scientific communication, for example) which ultimately set the framework for the act of communication both within and across cultural and linguistic boundaries. In fact, **the terminological aspect can be made to work most effectively only when**

²⁶ *A caça à terminologia tem encoberto o fato de que há algo entre os termos que é igualmente importante para uma tradução satisfatória. Em vez de se considerar o texto técnico como uma coleção de termos técnicos, ele deveria ser visto como um **texto** – isto é, como uma entidade escrita para servir a vários propósitos. A percepção desses propósitos não pode ser alcançada focando a terminologia e trabalhando sentença a sentença.*

²⁷ “Culture in scientific and technical communication: A dynamic view from a translation perspective” (Hatim, 1998: 2, unpublished) – o autor me informou, em 2008, que esse artigo foi transformado num capítulo de mesmo nome (capítulo 5) de seu próximo livro, ainda sem título.

seen within sound textual practice²⁸ (Hatim²⁷ *apud* Korning Zethsen, 1999: 73, grifo meu).

A autora termina seu artigo ressaltando que o sucesso da tradução técnica vai depender da capacidade do tradutor de perceber e levar em consideração todos os propósitos comunicativos do texto – e não apenas o informativo.

1.2.4 “Tradução técnica” neste trabalho

Com base no que destaquei no trabalho de alguns autores até aqui, é possível agora elaborar a visão de tradução a ser adotada na presente pesquisa, que aproveita elementos das várias teorias que fiz dialogar neste capítulo. Para mim, **atividade tradutória** pode ser entendida como um **ato comunicativo culturalmente condicionado cujo propósito é re-textualizar uma determinada mensagem numa segunda língua**. Dessa perspectiva, o resultado dessa atividade – a **tradução** propriamente – pode ser entendida como uma **re-textualização de um complexo de informações lingüísticas e extralingüísticas associadas a um texto de partida numa dada língua conforme são apreendidas e reformuladas numa segunda língua pelo tradutor**.

As condições de produção e recepção dessa mensagem re-textualizada serão determinadas, entre outras coisas: i) pelos **propósitos do iniciador** (aquele que quer comunicar a mensagem numa segunda língua/cultura); ii) pela **competência do tradutor** (que será responsável pela re-textualização dessa mensagem); iii) pelas **necessidades / expectativas do público-alvo** (os leitores/usuários potenciais desse texto); iv) pelo **gênero e tipologia textual** (meio de veiculação do texto, tipo de publicação / transmissão, tamanho da fonte, recursos gráficos, icônicos, etc.).

Uma tradução bem sucedida será aquela que, acima de tudo, cumprir seu propósito comunicativo, atendendo às expectativas e necessidades do público-alvo e do iniciador distando-se o menos possível da mensagem original. Com isso quero dizer que não concordo com a visão partilhada por autores como Vermeer (1987), Gutt (1991) e Pinchuck (1977) de que o texto original é mera “oferta de informação” pois, a

²⁸ (...) *aqueles aspectos negligenciados da prática textual (o discurso, o gênero e os textos reais que tipificam a comunicação científica, por exemplo) que, em última instância, estabelecem as condições de produção do ato comunicativo, seja dentro de uma mesma língua e cultura ou entre línguas e culturas distintas. Na verdade, o aspecto terminológico passará a funcionar apenas quando visto da perspectiva de uma prática textual consistente.*

exemplo do que defendem Azenha Jr. e Korning Zethsen, não considero o texto técnico como puramente informativo, ou como uma categoria totalmente diferente de qualquer outro tipo textual.

A noção de “function plus loyalty” (função mais lealdade) de Nord (1997: 125), definida como “a bilateral commitment on the part of the translator to both source and target sides of the translation interaction”²⁹ retrata melhor meu posicionamento. A *Skopostheorie* prevê que cada projeto de tradução resultará em “a free or faithful translation, or anything in between (...) depending on the purpose of the translation”³⁰ (*id. ibid.*: 29). Ou seja, a questão da fidelidade ao original vs. proximidade com o leitor poderia ser vista como uma “balança de pesos” (Costa, 2006: 28), especialmente em se tratando de textos culturalmente marcados (como é o caso da Culinária):

De um lado temos o texto-fonte (estratégias de equivalência formal, estrangeirização, opacidade, etc.), do outro a compreensão pelo público-alvo (equivalência dinâmica, domesticação, familiarização, etc.). A tarefa do tradutor é encontrar o equilíbrio entre os dois pratos cujos pesos irão variar de acordo com o gênero textual, a finalidade da tradução, as habilidades do tradutor, entre outros. (Costa, 2006: 28).

Mas, como a cultura e os fatores envolvidos na situação de comunicação – as condições de produção e recepção do texto original e da tradução – são expressas, se materializam lingüística e textualmente (sendo, portanto, passíveis de quantificação e análise)? O aspecto central dessa “materialização”, a meu ver, é o próprio texto – as palavras e como elas se organizam naquele espaço delimitado. O texto é, por assim dizer, a realidade palpável dessa interação cultural, o microcosmo que mimetiza, no embate das palavras, as relações de força, atração e repulsão dos aspectos extratextuais e culturais a que está condicionada a re-textualização da mensagem original. Texto e contexto são intimamente ligados e condicionam um ao outro. Ou, nas palavras de Hönig e Kussmaul (1982: 58 *apud* Azenha Jr., 1999: 33): *Todo texto pode ser entendido com a parte verbalizada de uma sociocultura.*

²⁹ um comprometimento bilateral por parte do tradutor com o lado fonte e o lado alvo da interação de tradução

³⁰ uma tradução livre ou fiel, ou qualquer ponto entre esses dois extremos, dependendo do propósito da tradução

Os textos são compostos de palavras. Mas as palavras, em si e individualmente (ou, em “estado de dicionário”), não detêm mais sentido que aquele que acumularam ao longo de sua vida discursiva – ou seja, seu sentido se (re)constrói no texto, à medida que cada uma delas negocia significados com as demais que ali estão (e com as que não estão), e com todo o entorno da situação de comunicação e da cultura. A **linguagem técnica** representa, nesse sentido, **uma forma típica de certas palavras se organizarem em determinadas situações de comunicação**. E é justamente essa forma típica de organização das palavras nas receitas culinárias que vai ser estudada neste trabalho. Mas antes, vejamos alguns conceitos fundamentais da tradução que serão importantes para esta pesquisa: unidade de tradução, equivalência, estratégias de tradução e naturalidade.

1.2.5 Alguns conceitos básicos da Tradução

1.2.5.1 Unidade de tradução

O conceito de “unidade de tradução” tem sido debatido e pesquisado empiricamente por vários estudiosos da tradução. A delimitação do que constitui tal unidade vai depender, em primeira instância, do posicionamento teórico do autor ante a tão aludida dicotomia *fidelidade vs. liberdade*, ou *tradução palavra-por-palavra vs. tradução sentido-por-sentido* e, conseqüentemente, a faceta privilegiada desse objeto de estudo: o texto original, o processo ou produto da tradução. Morfema, palavra, sintagma, frase, oração, parágrafo, texto, discurso e até mesmo cultura são alguns dos elementos passíveis de serem classificados como “unidade de tradução”.

Vinay e Darbelnet (1995 [1957]) são citados por vários autores como os primeiros a tratarem do tema no âmbito da tradução (Hatim e Munday, 2004; Alves *et al.*, 2003 [2000]; Mounin, 1963), e o fizeram do ponto de vista da Lingüística. Para Vinay e Darbelnet, a unidade de tradução pode ser definida como: “the smallest segment of the utterance whose signs are linked in such a way that they should not be translated individually”³¹ (*op. cit.*: 21). Os autores usam o termo “unidade de tradução”

³¹ *o menor segmento de um enunciado em que os signos estão relacionados de uma tal forma que não podem ser traduzidos individualmente*. Chamou-me a atenção a tradução dada para o mesmo trecho em Alves *et al.* ([2000] 2003), supostamente feita diretamente do original em francês, em que “whose signs are linked in such a way...” foi vertido como “cuja coesão de sinais seja tal...” – parece-me razoável pensar que a palavra “signs” (se a tradução foi feita corretamente do francês ao inglês) refere-se ao conceito saussureano de “signo” (significante + significado), não podendo ser traduzido neste contexto por “sinais”.

como sinônimo de “unidade de pensamento” e “unidade lexicológica” – “for us, these terms convey the same concept, but with emphasis on different points of view”³² (*id. ibid.*: 22). A razão disso é sua visão de tradução, fundamentada na teoria saussureana do signo:

Translators, let us remind ourselves, start from the meaning and carry out all translation procedures within the semantic field. They therefore need a unit which is not exclusively defined by formal criteria, since their work involves form only at the beginning and the end of their task. In this light, the unit that has to be identified is a unit of thought, taking into account that translators do not translate words, but ideas and feelings³³. (Vinay e Darbelnet, 1995 [1957]: 21).

Os autores afirmam, ainda, que os tradutores precisam de uma definição de unidade de tradução que leve em conta não apenas aspectos formais – tais como espaços em branco e sinais de pontuação. É possível identificar outras passagens que também tratam a unidade de tradução de uma perspectiva mais textual.

(...) after all, in written language utterances are divided into words by blank spaces and dictionaries are compiled on the principle of such units as words. But even in written language, the limits of a word are not always very clear^{34,35}. (Vinay e Darbelnet, 1995 [1957]: 20)

The translation of a word usually depends upon its context. A unit of translation provides a limited context: it forms a syntactic unit where one element determines the translation of the other. (...) On the other hand, the context is created by the usage, and it is unlikely that these words should recur in the same order with a different meaning association³⁶. (*op. cit.*: 27)

³² para nós, esses termos referem-se a um mesmo conceito, mas com ênfase em diferentes pontos de vista

³³ Tradutores, vale a pena lembrar, partem do sentido para realizar todos os procedimentos de tradução no âmbito do campo semântico. Portanto, necessitam uma unidade que não seja exclusivamente definida por critérios formais, já que seu trabalho envolve a forma apenas no início e no final de sua tarefa. Dessa perspectiva, a unidade que deve ser identificada é uma unidade de pensamento, considerando-se que os tradutores não traduzem palavras, mas idéias e sentimentos.

³⁴ afinal, na linguagem escrita os enunciados são divididos em palavras pelo espaço em branco e os dicionários são compilados partindo do princípio de que as palavras são unidades. Mas, mesmo na linguagem escrita, os limites de uma palavra nem sempre são muito claros.

³⁵ O autor exemplifica com palavras que ora aparecem com hífen, ora sem.

³⁶ A tradução de uma palavra geralmente depende de seu contexto. Uma unidade de tradução oferece um contexto limitado: ela forma uma unidade sintática em que um elemento determina a tradução do outro. (...) Por outro lado, o contexto é criado pelo uso, e é improvável que essas palavras possam ocorrer novamente na mesma ordem com uma associação de significado diferente.

Esses dois trechos de Vinay e Darbelnet ressaltam pontos importantes para a constituição de um conceito de “unidade de tradução” (algumas de suas proposições aproximam-se de conceitos usados em Lingüística de Corpus, como veremos mais adiante):

1. palavras – separadas por espaços em branco – não constituem unidades, necessariamente (e, portanto, não são os melhores itens para compor, por exemplo, a lista de entradas de um dicionário);
2. o sentido das palavras depende do contexto em que ocorrem, com o qual formam uma unidade;
3. a ocorrência de uma palavra num determinado contexto co-seleciona as demais palavras e influencia na determinação de seus sentidos;
4. combinações recorrentes de palavras específicas dispostas numa certa ordem têm um sentido razoavelmente estável.

Outros teóricos procuraram definir a “unidade de tradução” de outras perspectivas. Newmark (1988: 66) afirma que qualquer segmento lingüístico pode, em diferentes momentos e também simultaneamente, ser usado como unidade de tradução no decorrer da atividade tradutória, embora sugira que a maior parte deles restrinja-se ao nível da palavra – e, para ele, raramente no nível do parágrafo e jamais no nível do texto (Hatim e Munday, 2004: 25; Alves *et al.*, [2000] 2003: 31). Newmark considera que a “autoridade” de um texto é que vai, em última análise, definir o tamanho das unidades de tradução: *the more authoritative the text, the smaller the unit of translation*³⁷ (Newmark, 1988: 66).

Autores como Reiss e Vermeer (1984), Beaugrande (1978) e Snell-Hornby (1988), por outro lado, acreditam que a unidade de tradução deve ser mais abrangente: situam-na ao nível do texto, da intertextualidade e até mesmo da cultura, respectivamente – um posicionamento que reflete a influência de teorias de orientação funcional, discursiva e cultural nos Estudos da Tradução. Mas Hatim e Munday observam, apropriadamente:

(...) some theorists stress that the major unit must be the text itself, that it is impossible to translate well unless the significance of the whole text has been established. Yet practicing translators are often required to undertake a

³⁷ *quanto mais autoridade tiver o texto, menor será a unidade de tradução.*

translation without having had the time (or, in some instances, the opportunity) to read or access the whole of a lengthy text³⁸. (Hatim e Munday, 2004: 136).

Para as teorias centradas no texto de chegada (p. ex. Toury), uma unidade de tradução é uma unidade do texto traduzido que pode ser vinculada a (“mapped onto”) uma unidade no texto de partida. Toury (1986) demonstrou que as traduções produzidas por aprendizes inexperientes continham um número maior de segmentos, em geral restritos à palavra e, até mesmo, ao morfema, ao passo que a tradução feita por um aluno com experiência em tradução continha aproximadamente metade do número de unidades, e estas situavam-se no nível da frase e da oração – o que, em geral, confere ao texto de chegada um maior grau de aceitabilidade, segundo o autor (*apud* Malmkjaer, 1998: 286).

Resultados semelhantes foram obtidos por pesquisadores que estudaram o processo tradutório focando as operações mentais e cognitivas envolvidas na tradução (p. ex. Lörscher, 1991; Gerloff, 1988; Ericsson e Simon, 1984). Para a abordagem teórica processual, a unidade de tradução é um trecho do texto original no qual o tradutor foca sua atenção para então representá-lo na língua de chegada. Uma definição mais detalhada de unidade de tradução dessa perspectiva pode ser encontrada em Alves *et al.*:

UNIDADE DE TRADUÇÃO é um segmento do texto de partida, independente de tamanho e forma específicos, para o qual, em um dado momento, se dirige o foco de atenção do tradutor. Trata-se de um segmento em constante transformação que se modifica segundo as necessidades cognitivas e processuais do tradutor. A unidade de tradução pode ser considerada como a base cognitiva e o ponto de partida para todo o trabalho processual do tradutor. Suas características individuais de delimitação e sua extrema mutabilidade contribuem fundamentalmente para que os textos de chegada tenham formas individualizadas e diferenciadas. O foco de atenção e consciência é o fator direcionador e delimitador da unidade de tradução e é através dele que ela se torna momentaneamente perceptível. (Alves *et al.*, [2000] 2003: 38).

³⁸ (...) alguns teóricos enfatizam que a unidade de tradução principal deve ser o próprio texto, que é impossível traduzir bem a menos que o significado do texto como um todo tenha sido estabelecido. No entanto, tradutores profissionais não raro são instados a traduzir sem ter tido o tempo (ou, em alguns casos, a oportunidade) de ler ou manusear a totalidade de um texto longo.

Como é possível depreender da definição acima, a única forma de se delimitar a unidade de tradução para essa teoria é observando o processo por meio do qual tradutores traduzem textos (já que não é possível, ainda, saber exatamente quais são os processos mentais e cognitivos *per se* desses sujeitos durante tal processo) – o que é feito por meio de metodologias como os *Think-Aloud Protocols*. Resultados obtidos em pesquisas desenvolvidas usando essa metodologia demonstram, a exemplo do que Toury observou, que aprendizes tendem a focar a atenção em unidades de tradução menores, tais como morfemas e palavras, ao passo que tradutores experientes isolam e traduzem unidades maiores, ou “unidades de sentido”, que em geral correspondem a frases, orações e sentenças.

Em outra parte de seu texto, Alves *et al.* (2003 [2000]: 34) afirmam: “a delimitação das UTs [unidades de tradução] depende exclusivamente de cada um de nós e de nossa bagagem pessoal de conhecimentos”. Ainda que não possa discordar totalmente desse ponto de vista, ele impõe sérios problemas metodológicos a quem quiser pesquisar e descrever tais unidades, principalmente se considerarmos, com os autores, que as unidades de tradução são “um segmento em constante transformação que se modifica segundo as necessidades cognitivas e processuais do tradutor” (*op. cit.*: 38). Independentemente da estratégia usada pelo tradutor para segmentar o texto, e se ele é experiente ou não, me interessa saber se o resultado obtido – a tradução do texto em si – foi bem sucedido, ou seja, adequado à situação de comunicação em que se insere, ou não. Em caso afirmativo, acredito ser possível identificar **regularidades** na forma de segmentar tanto o texto de partida quanto o de chegada que, independente dos processos mentais que os condicionam (obviamente dependentes da experiência pessoal de cada um), realizam-se – isto é, tornam-se realidade – por meio de uma operação lingüístico-comunicativa e, portanto, ancorada no texto, em seu entorno e em seus componentes menores.

Nesse sentido, ainda que o texto de chegada tenha “formas individualizadas e diferenciadas”, conforme apontado por Alves *et al.* (no sentido de que não há como duas pessoas produzirem uma tradução idêntica), há certos critérios que permitem considerar uma tradução bem sucedida / adequada, ou não, especialmente no caso da tradução técnica. E esses critérios, como venho tentando explicitar desde o início deste trabalho, não se pautam apenas pela acuidade terminológica, mas também pela organização textual – que é culturalmente condicionada, como as demais instâncias do ato comunicativo.

Em outras palavras, os elementos que fazem com que uma tradução técnica seja considerada bem sucedida / adequada não podem ser identificados, a meu ver, por meio da simples comparação do texto de chegada com o texto de partida, mas com outros textos autênticos representativos do gênero e tipologia textual em questão na língua de chegada – afinal, são eles que ditam os parâmetros do que é bom e adequado, segundo os quais a qualidade do texto traduzido será julgada. Que unidades, nesses textos, são copiadas / repetidas / imitadas / mimetizadas na boa tradução? Essas serão, para nós, “unidades de tradução”, quer se situem no nível microtextual, macrotextual ou extralingüístico. E por restringir minha pesquisa a um domínio particular do saber, proponho denominá-las “Unidades de Tradução Especializada”, doravante UTEs.

Um dos objetivos deste trabalho é propor uma metodologia semi-automática de identificação dessas unidades. No entanto, não vou identificá-las por meio da observação do texto original, do processo de tradução, e nem mesmo de seu produto – o texto traduzido –, mas em textos semelhantes ao que o tradutor pretende produzir e que foram originalmente escritos na língua de chegada (o que farei segundo os pressupostos teóricos e metodológicos da Lingüística de Corpus). Somente depois de relacionar essas unidades a seu contexto de produção e recepção em cada língua é que poderei estabelecer possíveis relações de equivalência interlingual entre elas, o que farei recorrendo a teorias dos Estudos da Tradução sobre equivalência e estratégias de tradução, de que trato a seguir. No último passo do trabalho, será elaborada uma proposta de dicionário voltado para o tradutor que possa abrigar as UTEs extraídas do corpus.

1.2.5.2 Equivalência e estratégia de tradução

O conceito de equivalência é inerente à teoria da tradução. Ainda assim, é um dos mais controvertidos – diferentes abordagens dão-lhe importância e tratamento diversos. A própria palavra “equivalência” foi usada com acepções distintas por teóricos da área. Para Vinay e Darbelnet (1995 [1957]), por exemplo, “equivalência” é apenas um dos procedimentos de tradução – aquele em que uma situação do original é reproduzida usando-se palavras diferentes, recomendada pelos autores para a tradução de expressões idiomáticas (Alves et al., 2003 [2000]: 88).

Autores como Snell-Hornby (1988) e Gentzler (1993) defendem que a noção teórica de equivalência é irrelevante e até mesmo prejudicial para o estudo da tradução, já que se restringe, não raro, ao nível da palavra, e também porque geralmente assume que o texto é uma representação fiel da linguagem (Kenny, 1998: 77-80) – ou seja, uma equivalência entre textos representaria, necessariamente, uma correspondência entre dois sistemas lingüísticos.

Já para estudiosos como Catford (1965), Toury (1980, 1995), Koller (1995) e Pym (1992), a tradução em si só pode ser definida em termos das relações de equivalência que se estabelecem entre um texto de partida e um texto de chegada. O debate sobre a (im)possível “correspondência formal” entre sistemas lingüísticos abstratos (nível da *langue*), debatida inicialmente por Catford (1965: 27) e retomada posteriormente por vários autores, foi sendo aos poucos preterida pelos teóricos da tradução. A partir de Toury (1980: 24), a discussão sobre “traduzibilidade” passou a focar mais o que Catford chamou de “equivalência textual” – aquela que se dá entre dois textos reais, um de partida e um de chegada, no nível da *parole*.

À medida que os estudos da tradução passaram a considerar a equivalência como uma relação de correspondência entre textos produzidos em línguas distintas, e não mais entre sistemas lingüísticos e conceptuais distintos, seu estudo tornou-se mais empírico:

Once attention was focused on texts and utterances, many of the potential multiple meanings and functions of words and structures in a language system could be eliminated by reference to their **co-text and context**, making translation not only more tractable, but more realistic³⁹. (Kenny, 1998: 79, grifo meu).

Para Pym (1992), que vê a tradução como uma transação, a equivalência não é um fenômeno estritamente lingüístico – **a equivalência é uma entidade negociável**, e os tradutores são os responsáveis por essa negociação. Em consonância com essa visão da equivalência como algo não monolítico, mas relativo, estão os trabalhos de Toury (1980; 1995). Para este autor, a questão que se coloca no estudo comparativo de originais e suas traduções não é se existe ou não uma relação de equivalência entre

³⁹ *Uma vez que o foco passou a ser os textos e os enunciados, muitos dos vários sentidos e funções potenciais das palavras e estruturas de um dado sistema lingüístico puderam ser eliminados por meio da referência a seu co-texto e a seu contexto, o que tornou a tradução não apenas mais tratável, mas também mais realista.*

eles, mas de que *tipo* é essa equivalência e qual o seu *grau*. Toury, assim como Koller (1995: 196), acredita que a **relação de equivalência entre dois textos é circunstancial e historicamente determinada**:

Rather than being a single relationship, denoting a recurring type of invariant, it comes to refer to any relation which is found to have characterized translation under a specified set of circumstances⁴⁰. (Toury, 1995: 61)

A meu ver, a equivalência é variável porque os próprios conceitos o são, além de serem também circunstanciais, cultural e historicamente determinados – não existem *a priori*, mas constituem-se no texto e na/pela linguagem. Stolze (1982: 384 *apud* Azenha Jr., 1999) afirma, nesse sentido, que “(...) a equivalência interlingual é uma ilusão. As línguas não são sistemas rígidos, mas formações dinâmicas em constante mutação, combinação e evolução, além de estarem ligadas aos seus usuários”.

O que tem acontecido na prática terminológica bi- e multilíngüe, muitas vezes, é que, em vez de dedicarem-se à pesquisa de possíveis relações de equivalência entre entidades reais e palpáveis, encontráveis nos textos que de fato circulam entre os especialistas de uma dada área, grande parte dos terminólogos tem se dedicado a estabelecer equivalências entre sistemas lingüísticos e conceptuais (ainda que usem *corpora* eletrônicos para isso). Por acreditar que linguagem e pensamento são indissociáveis, acredito também, a exemplo dos teóricos da tradução das últimas décadas, que a equivalência é um fenômeno textual, vinculado a uma situação de comunicação específica, e não conceptual. Portanto, o estabelecimento de equivalentes interlinguais (inclusive os de valor terminológico) só pode ser feito por meio da observação da língua em uso – e, mais particularmente neste caso, do fenômeno tradutivo, que é quando esses sistemas lingüísticos e conceitos se digladiam diretamente no concreto da linguagem: o texto.

(...) não sobressai a tradução como campo exclusivo, mas, indiscutivelmente, como campo privilegiado de observação das diferenças lingüístico-culturais que se interpõem à comunicação social. Na diferença está o próprio paradoxo da tradução: a tradução se faz a partir da diferença, pois sem diferença não haveria porque traduzir (...) e, no entanto, almeja

⁴⁰ Em vez de ser uma relação singular, denotando um tipo recorrente de invariante, refere-se a qualquer relação que se descubra ter caracterizado a tradução sob um conjunto especificado de circunstâncias.

a identidade, busca fazer-se entender tal como o original se faz – ou se pretende fazer – entender. (Aubert, 2003a: 152).

Bem, se a tradução é o espaço privilegiado para observar as relações de equivalência entre textos e, por conseguinte, unidades de tradução especializada, por que não usar, então, textos paralelos (originais e respectivas traduções) na compilação de terminologia bilíngüe? Ora, se pudermos considerar que a relação de equivalência entre dois textos é negociada pelo tradutor de acordo com uma situação de comunicação específica, histórica e culturalmente determinada, é muito mais conveniente e verossímil observarmos os atributos dos textos originalmente escritos na língua de chegada que, num dado momento histórico, caracterizam a produção textual especializada entre os membros de um dado grupo profissional (atributos estes que o tradutor procurará mimetizar)⁴¹.

Além disso, *corpora* paralelos – quando se tem acesso a eles – refletem aquilo que já foi traduzido um dia (por vezes, inapropriadamente). Mas, novos textos, que exprimem novos conceitos, novas denominações, novos produtos, equipamentos, novas relações entre conceitos, etc. aparecem a todo momento. Como encontrar um equivalente na outra língua para algo que ainda não tenha sido (ou tenha sido pouco) traduzido, ou re-textualizado nessa outra língua? É neste ponto que o pesquisador que se propõe produzir um material voltado para o tradutor terá que abandonar o trabalho puramente descritivo e lançar mão de outros recursos – como a consulta a especialistas da área – para propor **estratégias de tradução** possíveis para esses casos, já que há poucas evidências na língua de chegada que suportem o estabelecimento de uma UTE equivalente.

Há na literatura vários modelos de estratégias de tradução. Todos eles propõem procedimentos a serem adotados pelo tradutor nos casos de não-equivalência ou de equivalência parcial. Barbosa (1990), após analisar os modelos de tradução de vários teóricos (tais como Nida, Catford, Vinay e Darbelnet, Newmark e Auber) apresenta sua “proposta de categorização dos procedimentos técnicos da tradução”, que se pauta pelo grau de divergência entre a Língua Original e a Língua Traduzida. A categorização subdivide-se em quatro eixos: i) convergência do sistema lingüístico, da realidade extralingüística e do estilo; ii) divergência do sistema lingüístico; iii) divergência do estilo e iv) divergência da realidade extralingüística. Treze procedimentos técnicos de

⁴¹ isso não impede, é claro, que *corpora* paralelos sejam usados na identificação de equivalentes *prima facie* (vide item 3.4.2 do Capítulo III).

tradução – que, segundo a autora, procuram responder à pergunta "como traduzir?" – distribuem-se entre esses eixos, conforme consta do quadro sinóptico apresentado por ela:

<i>Convergência do Sistema Lingüístico, do Estilo e da Realidade Extralingüística</i>	<i>Divergência do Sistema Lingüístico</i>	<i>Divergência do Estilo</i>	<i>Divergência da Realidade Extralingüística</i>
Tradução palavra-por-palavra			
Tradução literal			
	Transposição		
	Modulação		
	Equivalência		
		Omissão vs. Explicação	
		Compensação	
		Reconstrução	
		Melhorias	
			Transferência
			Transferência c/ Explicação
			Decalque
			Explicação
			Adaptação

Figura 2: Categorização dos Procedimentos Técnicos da Tradução (Barbosa, 1990: 93)

Para fins da presente pesquisa, o grande mérito do trabalho de Barbosa está na revalorização dos procedimentos usados pelos tradutores para resolver os casos de não-equivalência ("divergência extralingüística", nas palavras da autora), que são os que apresentam maior dificuldade para o tradutor e, ainda assim, são geralmente desprezados pela terminologia bilíngüe. Segundo Barbosa, autores como Vinay e Darbelnet julgavam o **empréstimo** e o **calque** ("decalque" para ela) como os procedimentos mais fáceis para o tradutor – simples cópia do original. A autora comenta, a esse respeito:

Não acredito, no entanto, que o empréstimo seja um procedimento tão fácil assim. Ele é usado quando há uma divergência tão grande entre as línguas, entre as realidades extralingüísticas expressas por meio delas, que falta a uma itens lexicais possuídos pela outra para designar objetos ou exprimir conceitos desconhecidos pela primeira, o que representa grande dificuldade para o tradutor e obstáculo para a tradução e compreensão do TLT [texto na língua traduzida] por seu leitor. (Barbosa, 1990: 81)

Observação semelhante é feita por Aubert (2003b). Segundo o autor, há duas linhas de pensamento no que diz respeito ao uso do empréstimo como procedimento tradutório. Uma, representada por autores como Wilss (1982) e Vinay e Darbelnet (1995 [1957]), vê no empréstimo uma ‘mera cópia’, fenômeno possível na tradução, em certas circunstâncias limitadas, mas de baixo interesse geral, exceto para o enriquecimento do vocabulário técnico das linguagens de especialidade (terminologia). A outra, defendida por autores como Venuti (1995), Barbosa (1990) e o próprio Aubert, propõe: “valorizar o empréstimo, juntamente com os decalques sintáticos, lexicais e semânticos, como ferramenta tradutória que dá vida a uma opção deliberada de política cultural”. Como reporta Aubert (2003b), alguns estudos em áreas das Ciências Humanas parecem indicar, no entanto, que esse recurso de tradução foi pouco utilizado até o final da década de 1990, quando começou o grande *boom* da globalização (Alves, 1983; Darin, 1986).

Curiosamente, o texto de Vinay e Darbelnet sobre o empréstimo menciona a Culinária. Para nós, isso é mais um indicativo da carga cultural do tema, que vai entrar em choque com a precisão pretendida pelas áreas técnicas e, especialmente, pelos textos instrutivos:

(...) in order to introduce the flavour of the SL [source language] culture into a translation, foreign terms may be used, e.g. such Russian words as ‘roubles’, ‘datches’ and ‘apparatchik’, ‘dollars’ and ‘party’ from American English, Mexican Spanish food names ‘tequila’ and ‘tortillas’, and so on⁴². (Vinay e Darbelnet, 1995 [1958]: 31)

Com a globalização, se por um lado nota-se um processo de acultramento decorrente da massificação do conhecimento, há também um movimento contrário de resposta, de que nações e minorias subjugadas estão aos poucos tomando ciência e explorando em seu próprio benefício – a valorização do local, do específico, das particularidades que distinguem esses grupos no grande mercado globalizado e aplainado culturalmente que é o mundo atualmente.

⁴² (...) para introduzir o sabor da cultura da língua fonte numa tradução, termos estrangeiros podem ser usados; por exemplo, palavras russas como ‘roubles’, ‘datches’ e ‘apparatchik’; ‘dollars’ e ‘party’ do inglês americano; nomes de comida do espanhol mexicano como ‘tequila’ e ‘tortillas’, etc.

A correlação das línguas funciona como uma importante estratégia de integração entre os povos que ampliaram suas fronteiras, posto que o multiculturalismo dos tempos atuais valoriza as identidades lingüísticas (...) (Krieger, 2006: 46).

Acredito que o empréstimo goza hoje na tradução de um status ainda mais valioso do que o reclamado por Barbosa no começo da década de 1990 – pode ser sinônimo de respeito à diferença, de alteridade e, principalmente, uma forma de o tradutor transportar o leitor diretamente para o que há de particular no universo cultural engendrado pelo texto original (já que parece restar cada vez menos disso no mundo globalizado): “(...) o empréstimo, pelo próprio efeito ‘estrangeirizador’ que provoca no espaço de recepção, nada oculta, antes, explicita a alteridade cultural”. (Aubert, 2003a). O empréstimo é usado em *Culinária*, por exemplo, para denotar *status*: *ciboulette* parece ser algo muito mais requintado e saboroso que “cebolinha francesa”. Como observa Alves (2006), é um fator importante não apenas para a constituição do vocabulário especializado das línguas – por exemplo, os empréstimos do árabe em língua portuguesa no domínio da alimentação, como: *arroz*, *alface*, *azeite*, e *cenoura* – mas em sua renovação.

Outro recurso intimamente ligado ao empréstimo, mas que também pode ser usado em isolado pelo tradutor para dizer a diferença é a **explicitação**:

A explicitação, por sua vez, embora possa ocorrer isoladamente, tende a assumir o papel de ferramenta auxiliar de outros procedimentos, notadamente do empréstimo. Como procedimento padrão, superficializa-se a mais das vezes como construção apositiva, podendo, no entanto, manifestar-se como nota de tradutor. (Aubert, 2003a)

Para nós, o paradigma de forma equivalente a compor um verbete de glossário voltado para o tradutor pode lançar mão dessas estratégias de tradução – uma sugestão de aposto e/ou nota de tradutor que possa ser reformulada, de acordo com a situação de comunicação específica, ou inserida diretamente no texto traduzido. E no caso da não-equivalência constatada (apesar da fragilidade histórica de uma constatação como essa no mundo globalizado), faz-se necessário fornecer informações conceptuais ao tradutor – definições, imagens, contexto de uso, etc. – para que este tenha condições de, de acordo com o que pede o momento, propor um equivalente adequado à situação de comunicação.

A falsa idéia passada pela globalização e pelos avanços da tecnologia de que é possível ter acesso a tudo em qualquer lugar a qualquer momento cai por terra ante a própria necessidade da tradução: a tradução existe porque existe a diferença – mesmo onde parece mais improvável –, e é por meio da tradução que essas particularidades lingüísticas e culturais podem ser entendidas e preservadas.

(...) nem mesmo nas referidas tecnologias de ponta, a ‘universalidade’ conceptual e, eventualmente, designativa constitui uma norma geral. Aqui, como em qualquer outro tipo de discurso, variam as visões-de-mundo, variam os referentes: a fibra de vidro, embora de idêntica composição físico-química, não tem as mesmas aplicações num país tropical e em um país subpolar; o aparelho eletrônico de última geração não será configurado da mesma maneira em um país com tensão elétrica uniforme e constante e em outro com tensão elétrica variável e inconstante. (Aubert, 1992: 83).

Mas não é tarefa fácil caracterizar essas diferenças. Schmitt (1989 *apud* Azenha Jr., 1999: 79-83), que trabalha da perspectiva das áreas técnicas, elaborou uma tipologia para tentar classificar o que ele chamou de “incongruências conceptuais interculturalmente condicionantes”, quais sejam:

1. Diferenças nas hierarquias conceptuais;
2. Diferenças na construção de produtos condicionadas por normas e leis específicas de cada cultura;
3. Diferenças na construção de produtos condicionadas por diferenças climáticas;
4. Recomendações de produtos, específicas de cada cultura;
5. Textos padronizados, próprios de cada cultura;
6. Orações declarativas próprias de cada cultura;
7. Exemplos específicos de determinada cultura;
8. Ilustrações específicas de cada cultura;
9. Métodos de medição específicos de cada cultura;
10. Falta de padronização terminológica entre empresas do mesmo ramo, e mesmo no interior de uma mesma empresa.

É possível que adaptações precisem ser feitas a este modelo caso ele seja aplicado à Culinária, pois é visível que o autor se baseou em áreas específicas para

elaborá-la. Mas discutirei essas alterações apenas no item 2.4 do Capítulo II, após caracterizar o tipo textual “receita”.

1.2.5.3 Naturalidade

O conceito de *naturalidade* é de suma importância para este trabalho, pois é o que torna evidente, na minha opinião, a necessidade de se adotar a Linguística de Corpus como a perspectiva teórica e metodológica mais adequada para o estudo tanto da Tradução Técnica quanto da Terminologia Multilíngüe.

Há vinte e cinco anos, Pawley and Syder (1983) publicaram um trabalho demonstrando que era possível, por meio da aplicação de regras gramaticais a um léxico, produzir sentenças “gramaticalmente corretas”, mas que dificilmente (ou mesmo nunca) seriam produzidas por um falante nativo. O aprendiz de língua estrangeira, de acordo com esses autores, costuma ver a língua dessa forma: constrói frases que estão corretas em teoria, mas que soam nada naturais para o falante nativo da língua que está aprendendo (Kjellmer, 1992). O tradutor inexperiente, como atestam vários autores, que corroborei em trabalho anterior (Teixeira, 2004), costuma agir da mesma forma ao traduzir.

Tagnin, em seu livro de 2005, *O jeito que a gente diz*⁴³, ilustra bem esse comportamento descrevendo com humor uma situação vivida por ela mesma ao desembarcar pela primeira vez em terras estrangeiras:

Quando fui aos Estados Unidos pela primeira vez, em 1961, já estava estudando inglês há seis anos, tendo sempre sido das melhores alunas. Acreditava – santa ingenuidade! – ter um bom domínio da língua inglesa. Entretanto, num de meus primeiros encontros, expressei-me de forma a fazer com que meu interlocutor me olhasse como se eu houvesse cometido um erro gramatical gravíssimo. Perguntei-lhe:

— Não está certo o que eu disse?

— Está - respondeu ele -, só que não é desse jeito que a gente diz. (Tagnin, 2005: 12)

Uma situação como essa descreve o que Fillmore (1979) chamou de “falante ingênuo”, que Tagnin caracteriza, baseada nesse autor, como aquele que:

⁴³ Edição revisada e ampliada de seu *Expressões idiomáticas e convencionais*, publicado pela série Princípios da Ática, em 1989.

- i. desconhece os lexemas idiomáticos de uma língua;
- ii. não conhece as frases idiomáticas da língua;
- iii. não conhece as combinações lexicais que não se baseiam necessariamente nas relações de significado;
- iv. não tem capacidade de julgar a adequação de expressões fixas a certos tipos de situação;
- v. não conhece as imagens metafóricas da língua;
- vi. não entende os atos de comunicação indireta;
- vii. desconhece as convenções das estruturas de diversos textos.

Tagnin (2002a, 2002b) aplica acertadamente a noção de falante ingênuo de Fillmore ao âmbito da tradução, cunhando o termo “tradutor ingênuo”:

Basicamente, a ingenuidade do tradutor se configura numa compreensão composicional do significado e numa falta de consciência do quanto **uma língua é constituída dessas partes pré-fabricadas**.

A ingenuidade do tradutor pode transparecer tanto na sua habilidade de compreensão, quanto na de produção. (...)

Por estranho que pareça, mesmo como falante nativo da língua alvo, o tradutor pode ter problemas no nível da produção para conseguir **soluções naturais**, caso se atenha tanto ao texto de partida a ponto de não perceber que, **entre formas igualmente gramaticais, uma delas é de uso mais corrente**. Em outras palavras, pode não se dar conta de que, dentro de uma gama de formas gramaticalmente *possíveis*, há certas formas que têm uma **probabilidade maior de ocorrerem**. (Tagnin, 2002a: 192-193, itálicos da autora, grifo meu)

As observações de Tagnin encontram apoio em outros autores que estudam a tradução e, em particular, a tradução técnica:

Documents must speak ‘the language’ of the target audience and should resemble other texts produced within that particular language community and subject domain (...). These considerations frequently require that translators move beyond *merely correct* strategies in terms of lexical and grammatical

content in order to account for *stylistically appropriate* solutions.⁴⁴ (Wright, 1993: 70, itálicos da autora)

(...) even when translating into their mother tongue, translators do not always have the specialized linguistic knowledge (e.g. vocabulary, syntax, phraseology) required to express the subject matter in the appropriate way.⁴⁵ (Bowker, 2000: 185)

É possível depreender das citações acima, como já vimos em vários momentos deste capítulo, que esse “jeito que a gente diz”, no caso do texto técnico, não pode ser caracterizado apenas pelo uso correto de uma terminologia especializada. Esse ‘algo mais’ tem merecido a atenção de vários lingüistas, entre eles Fillmore, para quem a língua possui um *phrasicon*, ou seja, diversas expressões que funcionam como **blocos pré-fabricados** e “devem ser vistas tanto como itens lexicais quanto como possuindo uma estrutura gramatical que se situa num nível além da dimensão da palavra”⁴⁶ (Fillmore, 1978: 149). Ou, nas palavras de Altenberg e Eeg Olofsson (1990: 2), as expressões pré-frabricadas são como “uma espécie de ‘piloto automático’ que o falante pode acionar para que lhe sobre mais tempo para os aspectos criativos e sociais do processo de fala”⁴⁷. São essas expressões que vão garantir a naturalidade no texto traduzido (vide também item 3.3.3 do Capítulo III).

Vejamos, a seguir, em que medida os dicionários especializados têm contemplado as questões tão caras ao tradutor discutidas até aqui.

1.3 Tradução técnica e dicionários

A preocupação com a tradução técnica esteve sempre presente, de alguma forma, nas reflexões sobre terminologia, especialmente depois da virada lingüístico-computacional ocorrida na disciplina a partir do início da década de 1990. E a recíproca é verdadeira – como procurei demonstrar anteriormente neste capítulo, a terminologia

⁴⁴ *Documentos devem falar “a língua” do público-alvo, e devem assemelhar-se a outros textos produzidos no âmbito daquela comunidade lingüística e subárea em particular (...). Essas considerações em geral requerem que o tradutor vá além das estratégias que buscam o meramente correto em termos lexicais e gramaticais de modo a encontrar soluções que sejam estilisticamente apropriadas.*

⁴⁵ *(...) mesmo ao traduzir para a sua língua materna, tradutores nem sempre possuem o conhecimento lingüístico especializado (p. ex. vocabulário, sintaxe, fraseologia) necessário para expressar o tema tratado de maneira adequada.*

⁴⁶ *have to be viewed both as lexical items and as having grammatical structure on a higher level than that of word dimension.*

⁴⁷ *a kind of ‘autopilot’ which the speaker can switch on to gain time for the creative and social aspects of the speech process*

tem sido, tradicionalmente, um dos focos principais de atenção da tradução técnica⁴⁸. Com efeito, como observa Van Hoof, é de se pensar se seria mesmo possível, em alguns casos, divisar a atividade tradutória da dicionarística:

Os leigos associam os tradutores aos dicionários. Afinal, o dicionário não é um instrumento essencial para o trabalho do tradutor? Os dicionários de mais de uma língua são a primeira coisa que vêm à mente. No entanto, não são a única ferramenta usada pelo tradutor, cujo arsenal está constituído também pelos grandes dicionários de uma só língua (...), além de obras enciclopédicas e especializadas. Vale perguntar, assim, em que medida os tradutores estiveram empenhados em construir esse arsenal? Que papel desempenharam nesse particular, através dos tempos, e que papel continuam a ter? Quem chegou em primeiro lugar: o tradutor ou o lexicógrafo? (Van Hoof, 1998: 241)

Dicionários compilados por tradutores, técnicos ou não, e parcerias entre tradutores e terminólogos (como os projetos *Termium* do Canadá e *EuroDicAutom* da União Européia) ilustram o envolvimento profundo entre tradução técnica e terminologia ao longo da história da Humanidade (Barros, 2007: 18). No entanto, há pouco registro e, menos ainda, teorização sobre essa interação e suas práticas.

Vários terminólogos mencionam a interface entre tradução e terminologia na literatura geral da área – por exemplo: Barros (2007 e 2004), Krieger & Finatto (2004), Krieger e Maciel (2001) e Aubert (1992 e 1996), no Brasil. Com exceção do texto de Aubert, publicado antes, todas as autoras mencionam os textos publicados em 1999 por Cabré, um dos maiores expoentes da Terminologia atualmente, em coletânea de artigos intitulada *La Terminología. Representación y comunicación*. Três capítulos da obra abordam diretamente a tradução técnica do ponto de vista da Terminologia: “Sobre la formació del traductor en terminologia” (1999a: 363-369), “Las fuentes terminológicas para la traducción” (1999a: 203-229) e “Traducción y terminología: un espacio de encuentro ineludible” (1999a: 177-201). Neste último, consta o trecho muitas vezes citado em artigos e apresentações que falam dessa relação entre as áreas:

Ningún especialista mínimamente informado en lingüística aplicada pone hoy día en cuestión que entre la traducción especializada y la terminología existe una relación evidente e

⁴⁸ ainda que a tradução técnica tenha começado a fazer parte das preocupações da teoria geral da tradução só muito recentemente.

inevitable, pero sin embargo se ha estudiado muy poco las características y motivaciones de esta relación y menos aún se han establecido sus límites. (Cabré, 1999a: 177)

Na literatura publicada desde então sobre o tema, os trabalhos que mais contribuem para entender as necessidades terminográficas específicas do tradutor, em sua maioria, são realizados por pesquisadores ligados ao estudo, à prática ou ao ensino da tradução técnica e de linguagens para fins específicos. Esses pesquisadores perceberam, no dia-a-dia de seu trabalho, as deficiências de grande parte dos materiais de consulta disponíveis atualmente e propuseram soluções para este e outros problemas que surgem da interação entre tradução e terminologia (p. ex.: Bowker, 1996; Pearson, 1998; Fuentes Morán e García Palacios, 2002; Pérez, 2002; Maia, 2002; Avance de Souza, 2003; Gómez e Vargas, 2004; Zanatta, 2004; Gómez, 2005; Arribas Baño, 2005; Salgado, 2006; Carvalho, 2007; Fromm, 2008). Como essas propostas envolvem de alguma forma o uso da Linguística de Corpus, serão abordados no Capítulo III, em especial no item 3.4.3. Vejamos, a seguir, como essa interface entre Tradução técnica e dicionários permeou a teoria e a prática da Terminologia nas últimas duas décadas.

1.3.1 Tradução técnica e terminologia: caminhos entrecruzados

A estreita relação entre tradução técnica e terminologia vem sendo percebida e explorada não só por teóricos da tradução, como pontuei até aqui, mas também por terminólogos, especialmente aqueles que se dedicam à pesquisa das linguagens de especialidade em contextos bi- e/ou multilíngües.

A tradução faz parte da própria gênese da Terminologia moderna. Ainda que o aparecimento da palavra “terminologia” remonte ao século XVII, quando passa a comparecer como verbete nos dicionários clássicos europeus, a disciplina inaugurou-se apenas na primeira metade do século XX, com base nos estudos realizados pelo engenheiro austríaco Eugene Wüster (1898-1977), que postulou como objetivo principal da Terminologia evitar a ambigüidade na comunicação interlingual entre profissionais e especialistas (Cabré, 2003: 179). Na esteira do que vinha acontecendo com as ciências em geral nos séculos XVIII e XIX (como o estabelecimento de taxonomias, sistemas classificatórios e nomenclaturas científicas, bem como a crescente internacionalização dos saberes), a teoria de Wüster se propôs a

sistematizar os conhecimentos científico-tecnológicos da época para que fossem transmitidos adequadamente de uma nação a outra e, portanto, de uma língua a outra.

De lá para cá, essa necessidade só fez ampliar-se com a globalização e o desenvolvimento científico acelerados, aumentando ainda mais a demanda por traduções técnicas e, conseqüentemente, por materiais de apoio adequados para sua boa prática:

Neste contexto de alargamento de fronteiras e de grande ampliação de intercâmbios, as línguas passaram a entrar mais fortemente em contato, exigindo novas competências lingüísticas, em que se inclui o domínio dos termos técnicos. Junto a essas novas necessidades encontra-se a crescente demanda pelas traduções técnicas, as quais necessitam transpor adequadamente as terminologias de uma língua para outra. (Krieger e Finatto, 2004: 18)

O desenvolvimento da Terminologia foi também influenciado, em grande medida, pelos avanços da Informática que, a partir dos anos 80, passou a permitir o processamento de uma quantidade cada vez maior de dados, favorecendo a criação de fichas terminológicas, bancos de dados, *corpora* e dicionários informatizados.

No entanto, se a principal preocupação da Teoria Geral da Terminologia (TGT) de Wüster era normatizar e controlar o uso de termos técnico-científicos entre especialistas (a criação do comitê de Terminologia – Comitê Técnico 37 – junto à ISO é prova disso), nas últimas décadas, com a divulgação da ciência em grande escala, os campos de atuação dessa disciplina se alargaram, assim como cresceu o número de usuários diretos e indiretos que precisam da terminologia no seu dia-a-dia:

Os resultados das pesquisas terminológicas têm inúmeras aplicações e a cooperação se dá em diversas áreas, tais como a tradução especializada, a documentação, o jornalismo científico, as ciências sociais, o ensino de línguas, o ensino de disciplinas técnicas e científicas. A terminologia fornece ainda dados para atividades de planejamento lingüístico e normalização terminológica, entre outras. (Barros, 2006: 22)

Essa expansão das áreas técnicas e a consideração dos novos usuários dessas terminologias tornaram necessária uma revisão dos pressupostos iniciais formulados por Wüster em sua tese de doutoramento, defendida em 1931. É inegável a contribuição do autor para o estabelecimento das bases epistemológicas, teóricas e

práticas que instituíram a Terminologia como disciplina. Mas seu enfoque voltado para a padronização e, em especial, sua insistência na noção de que os conceitos – o verdadeiro objeto de estudo da Terminologia para ele – são universais, e existem independentemente dos signos (que seriam meras etiquetas a eles associadas), começou a receber várias críticas a partir da última década do século passado. Com isso surgiram novas abordagens que, cada uma a sua maneira, ampliaram o objeto de estudo da Terminologia para abranger, a exemplo do que ocorreu nas teorias de tradução, não apenas o microuniverso do termo, mas os diferentes aspectos envolvidos na comunicação especializada como um todo – razão pela qual a Lingüística passou a desempenhar um papel mais central no panorama teórico multidisciplinar da Terminologia.

A Lingüística responde, portanto, pelo grande redirecionamento dos estudos terminológicos, ocorrido, especialmente, a partir dos anos 90 do século XX. Mas, a nova visão epistemológica foi também impulsionada por pesquisadores de formação filosófica e tradutológica. Há, igualmente, significativas contribuições advindas da inteligência artificial que têm motivado a Terminologia a avançar numa linha lingüística (Krieger e Finatto, 2004: 38).

No Brasil, os estudos terminológicos foram introduzidos no currículo de Lingüística de alguns centros de excelência nacionais (como a Universidade de São Paulo, a Universidade de Brasília e a Universidade Federal do Rio Grande do Sul) no início dos anos oitenta (Barros, 2004: 36). De lá para cá, a pesquisa em Terminologia, acompanhando o desenvolvimento mundial da comunicação especializada, passou a ser feita em várias universidades brasileiras e, com isso, muito evoluiu. Um exemplo disso é o fato de a edição especial da revista *Ciência e Cultura* do segundo trimestre de 2006, publicada pela SBPC e pela Imprensa Oficial, ter dedicado à Terminologia a capa e todo o conteúdo das 33 páginas de seu Núcleo Temático. O caderno conta com artigos de alguns dos maiores especialistas brasileiros na área, enfocando diferentes perspectivas aplicadas e teóricas, para em seguida listar os 23 grupos – restritos apenas àqueles registrados no diretório dos *Grupos de pesquisa no Brasil* do CNPq – que trabalhavam com Terminologia à época (atualmente⁴⁹, eles já somam 35).

A evolução se faz sentir também em congressos e eventos especializados da área de Letras – a Terminologia figura em todos eles como área de interesse para

⁴⁹ 29/4/2008

proposição de apresentações individuais, grupos de trabalho e, não raro, como tema central de discussão, alvo de plenárias e simpósios. Além disso, editoras brasileiras lançaram algumas importantes obras teóricas sobre o tema nos últimos 10 anos, entre livros introdutórios, números temáticos de revistas especializadas e coletâneas de artigos especializados.

Outro progresso visível pode ser observado no número de glossários técnicos (e entre eles, os bilíngües) produzidos recentemente. Muitos, de divulgação mais restrita, são resultado de trabalhos de conclusão de curso de alunos de especialização em tradução, ou são fruto de mestrados e doutorados das áreas de Letras e Linguística de todo o país. Há também coletâneas recém-criadas por editoras técnicas, ou mais voltadas para o ensino de línguas estrangeiras, como as séries *Mil e um Termos*, da SBS, e *Easy Way*, da Disal.

É fato que a Terminologia evoluiu muito no Brasil e no mundo desde o seu estabelecimento como disciplina, mas a nós interessa saber em que pé estão as reflexões sobre a gestão de terminologia voltada para o tradutor – se e *como* estão sendo pensados os materiais de referência bi- e multilíngües, tão necessários à prática da tradução técnica. Assim como procurei demonstrar anteriormente a importância crescente da tradução técnica dentro das preocupações teóricas e práticas dos Estudos da Tradução, a meu ver, o tradutor técnico, público-alvo preferencial e com necessidades informacionais bem particulares, também deveria estar no cerne das preocupações teóricas e práticas da Terminologia bi- e multilíngüe atual.

1.3.2 Wüster revisitado: uma nova visão do texto técnico

Felber (1984) e Laurén e Picht (1993), ao traçarem um histórico do surgimento da Terminologia como disciplina e descreverem seus fundamentos científicos basilares, distinguem três “escolas” de pensamento de onde se originaram os conceitos teóricos e práticos que fomentaram seu desenvolvimento enquanto disciplina: a Escola de Viena (Wüster, Felber), a Escola de Praga (Drodz, Havranek, Horecky, Roudný, Kocourek) e a Escola Soviética (Lotte, Rondeau, Drezen, Vinokur, Marusenko)⁵⁰. Reconhecem também os importantes avanços obtidos em centros de pesquisa do Canadá (p. ex. Meyer), dos países nórdicos (p. ex. Temmerman) e do Reino Unido (p. ex. Sager).

⁵⁰ para um maior aprofundamento, vide Temmerman (2000: 18-22) e Barros (2004: 49-56).

Auger (1988 *apud* Krieger e Finatto, 2004: 28 e Barros, 2004: 46), por seu turno, distingue três tendências principais segundo as quais a Terminologia se desenvolveu desde seu surgimento formal:

- uma tendência de orientação *lingüístico-terminológica*, que objetiva promover a univocidade na comunicação entre especialistas por meio da criação de normas e diretrizes a serem seguidas pelos usuários das terminologias;
- uma tendência preocupada com a *planificação lingüística*, que visa aparelhar as línguas regionais em situações de bilingüismo ou coexistência para valorizá-las e preservá-las;
- uma tendência voltada para a *tradução*, muito presente nos países com mais de uma língua oficial, que se concentra no estudo e na produção de obras de referência bi e/ou multilíngües.

Essas tendências podem ser observadas, em maior ou menor grau, nas várias correntes teóricas que, a partir de 1990, passaram a reeditar os paradigmas de trabalho e reflexão sobre a Terminologia no cenário mundial. Essa revisão dos pressupostos teóricos da TGT sugeriu mudanças também nas metodologias de compilação e edição de dicionários terminológicos – ferramenta de trabalho indispensável ao tradutor técnico.

A **Teoria Sociocognitiva da Terminologia** (TST) propõe uma reformulação da TGT considerando as possíveis contribuições da semântica cognitiva para a teoria e a prática da Terminologia. Rita Temmerman é o nome mais representativo dessa linha teórica, que tem como um dos marcos iniciais sua tese de doutorado, defendida em 1998 e depois transformada em livro e publicada pela John Benjamins, em 2000, sob o título *Towards new ways of terminology description: The sociocognitive approach*.

Ao criticar a TGT, Temmerman questiona os princípios basilares propostos por Wüster partindo do pressuposto de que, contrariamente ao que ele e outros terminólogos da chamada Escola de Viena (como Felber e Budin) pregam, “a terminologia só pode ser estudada no discurso” (Temmerman, 2004 [2001]: 37). Suas críticas concentram-se em dois pontos principais da teoria “tradicional”: i) a primazia do conceito sobre o signo lingüístico, e as conseqüências disso para o modelo de análise do significado (*op. cit.*: 43); ii) a preocupação normatizadora, que preconiza a relação biunívoca entre conceito e termo e despreza as situações comunicativas e cognitivas de uso da linguagem especializada (*op. cit.*: 33).

Para a autora, o ponto de partida da Terminologia não deve ser mais o conceito, unidade abstrata e (supostamente) definível, mas o termo – unidade do discurso. Afirmar, apoiada em autores como Fillmore (1985), Fillmore & Atkins (1992) e Lakoff (1987), que as pessoas não usam conceitos para entender o mundo, mas sistemas cognitivos, em que “unidades de interpretação” – que corresponderiam, em Terminologia, aos termos encontrados nos discursos das áreas especializadas – encontram-se estruturadas prototipicamente e relacionadas entre si. A partir da análise do discurso especializado, o terminólogo teria acesso a essas “categorias” que, segundo ela, não podem ser claramente delineadas, pois é como se cada uma fosse *“um bloco de conhecimento que possui um núcleo e uma estrutura, mas que existe num processo de reformulação contínua e está, portanto, em constante transformação”* (op. cit.: 37, itálico da autora).

Essa nova aproximação proposta por Temmerman – segundo ela, mais preocupada “com a descrição autêntica do significado dos termos encontrados em arquivos textuais” (op. cit.: 32) – traz consequências para a Teoria da Terminologia como um todo, exigindo um reexame dos fundamentos teóricos e práticos postulados por Wüster. Por exemplo: os terminógrafos partiriam não mais dos conceitos para descrever a terminologia de uma dada área – segundo ela, uma abordagem “artificial” imposta pela TGT –, mas das unidades de interpretação contidas em textos especializados. Somente após a compilação de um *corpus* textual, que deve ser “selecionado a partir das **necessidades do grupo de usuários potencial** da terminologia” (op. cit.: 45, grifo meu), é feito um levantamento dos termos nele contidos para posterior coleta de informações.

A TST, além de reforçar a necessidade de interação entre Terminologia e sistemas computadorizados de gestão terminológica, enfatiza três outros aspectos da compilação de dicionários técnicos que parecem contribuir para as reflexões teóricas e metodológicas do presente trabalho: i) a consideração do **público-alvo** na seleção e descrição das unidades de interpretação que comporão a obra terminográfica; ii) o uso mandatório de dados provenientes de *corpora* textuais, contendo **discursos reais**, para a pesquisa terminológica; iii) a proposta de organização da terminologia das áreas de especialidade com base em **estruturas prototípicas** e modelos cognitivos⁵¹.

⁵¹ embora não tenhamos como objetivo descrever as relações de conteúdo dos conceitos da área Culinária neste trabalho (o dicionário resultante desta pesquisa não será de natureza definitiva), o

A importância de considerar a terminologia em uso é também o ponto central de crítica à TGT sustentado pela **Socioterminologia**, corrente teórica que dedica-se à preservação da identidade lingüística e cultural das terminologias, feita por meio da planificação lingüística⁵², e tem na variação terminológica seu foco de estudo principal. Representada por autores como Boulanger, Gambier e Gaudin, essa corrente teórica desenvolveu-se bastante nos países em situação de bilingüismo, como o Canadá⁵³, impulsionada principalmente pela demanda tradutória. Surgiu como uma reação à hipernormalização, para atenuar os efeitos prescritivos e normativos que acabavam por desvincular a terminologia das situações lingüísticas particulares de cada cultura (Faulstich, 2006: 29). Tem como marco inicial a tese de doutorado de Gaudin (1993), em que o autor postula que os termos devem ser estudados “no âmbito das práticas lingüísticas e sociais concretas dos homens que empregam tais termos” (Gaudin, 1993: 216 *apud* Faulstich, *id. ibid.*).

A Socioterminologia reúne elementos de ciências como a sociolingüística, a sociologia, a semântica cognitiva e as políticas lingüísticas. Opõe-se a vários pressupostos da Teoria Geral da Terminologia (TGT). Primeiramente, insiste que o estudo dos termos não pode ser feito “*in vitro*”, isto é, isolando-se os termos de seus contextos e das situações sociais em que ocorrem, mas somente “*in vivo*”, ancorando a comunicação especializada a seu lugar social. Reconhece, ao contrário da TGT, a existência de variação lingüística nos discursos especializados. A unidade terminológica é considerada como um signo lingüístico, um conteúdo indissociável de sua expressão, e não como um conceito cuja existência tem primazia sobre a forma, como quer a TGT. Por considerar a dimensão histórica da comunicação especializada, a Socioterminologia dedica-se ainda a estudar fenômenos desprezados pela TGT, como a banalização do discurso técnico-científico (Barros, 2006: 69-70).

A reformulação da Terminologia proposta pela Socioterminologia coloca sob escrutínio os materiais terminológicos produzidos seguindo-se os pressupostos da TGT, “propondo que o artificialismo do ideal normalizador seja suplantado pelo exame do contexto de produção dos léxicos especializados” (Krieger e Finatto, 2004: 35). Ou

modelo sugerido pela autora parece ser o mais adequado para representar a relação entre conhecimentos nas áreas especializadas, como também defende Lara (1999).

⁵² intervenção na língua no sentido de criar novas terminologias para preencher lacunas de vocabulário, ou para encorajar seu crescimento e desenvolvimento.

⁵³ influenciou também a formação da Terminologia no Brasil, já que professores da área de Lingüística, agraciados com bolsas de estudo da Associação Brasileira de Estudos Canadenses, voltaram de lá trazendo na bagagem as novidades teóricas e metodológicas desse campo de estudos (Finatto, 2001).

seja, a terminologia, como estou procurando argumentar neste trabalho, deve ser considerada em seu **espaço de interação social** – aspecto indispensável também para o estudo da tradução técnica, conforme discutirei com mais detalhe no Capítulo II, quando tratar da Culinária.

De todas as propostas críticas ao trabalho de Wüster, a **Teoria Comunicativa da Terminologia** (TCT) foi a que melhor se estruturou, construindo um arcabouço teórico e prático ao longo das últimas duas décadas que conquistou muitos adeptos em todo o mundo, inclusive no Brasil. Uma série de artigos publicados no final da década de 1990 por Maria Teresa Cabré, alguns em co-autoria com colegas seus do IULA (Instituto de Lingüística Aplicada) da Universidade Pompeu Fabra, Barcelona, são considerados por vários autores como o marco inicial dessa linha de pesquisa. Segundo Krieger e Finatto (2004: 35), a esse grupo deve-se o pioneirismo no debate sistemático e crítico à Teoria Geral da Terminologia (TGT) de Wüster.

A crítica de Cabré (1999a) à TGT concentra-se em quatro pontos que a autora considera fundamentais para entender o léxico especializado: i) sua poliedricidade (denominativa, cognitiva e funcional); ii) sua dupla função (representativa e comunicativa); iii) a definição de seus elementos operantes (a linguagem é real, e não ideal; a comunicação deve ser considerada *in vivo*, e não *in vitro*); iv) a diversidade de suas aplicações (de acordo com as características pragmáticas da comunicação). A autora critica também a aplicabilidade das metodologias propostas por Wüster, voltadas para a padronização da comunicação especializada. Segundo a pesquisadora, elas não podem dar conta das diversas necessidades terminológicas e da dinâmica constante dos âmbitos especializados, além de não incorporarem as novas tecnologias da informação.

Cabré introduz a perspectiva lingüística nos estudos terminológicos ao enfatizar que uma das características do termo é pertencer às línguas naturais. Nenhuma unidade lexical, para a TCT, pode ser considerada *a priori* como termo (ou, na terminologia de Cabré, “unidade terminológica” – UT), pois assume esse valor em função de seu uso num contexto comunicativo especializado. Essa constatação, para a autora, ressalta na TGT “la contradicción que suponía defender la univocidad y la monosemia de los términos sin distinción de situaciones de comunicación”. (Cabré, 1999a: 125-6).

Para dar conta das muitas facetas do termo (que é “poliédrico” por natureza), a autora cria um modelo que representa o marco teórico de sua proposta. Inspirada no modelo tridimensional de Sager (1990: 13), a “Teoria das portas”, como foi batizada por Cabré, prevê que a unidade terminológica, objeto de estudo da Terminologia, pode ser comparada a uma casa com múltiplas portas. As teorias distintas segundo as quais essa unidade terminológica pode ser explicada são como chaves – cada qual abre uma porta, e cada porta dá acesso por um cômodo distinto da casa, oferecendo uma perspectiva diversa: lingüística (teoria da linguagem, do signo), cognitiva (filosofia, psicologia, teoria do conhecimento) e comunicativa. A TCT é, portanto, uma teoria **lingüística** de base **cognitiva**⁵⁴ e propósito **comunicativo**. Nas palavras da autora:

(...) una macroteoría lingüística constituida por tres teorías: una teoría de la gramática que incluya la variación dialectal y funcional; una teoría de la adquisición que explique en una sola propuesta cómo se adquiere el conocimiento y dé cuenta de las correlaciones, identidades y diferencias entre las características y el proceso de adquisición del conocimiento general y el especializado en toda su diversidad funcional; y una tercera teoría de la actuación que recoja en un solo modelo el uso general y el especializado en toda su variada amplitud (temática, perspectiva, nivel de especialización, propósito comunicativo, propósito funcional, tipo de texto, tipo de discurso, etc.). (Cabré, 1999a: 120)

Outra característica da proposta da TCT, como método científico, é seu caráter indutivo. Ao contrário do que postula a TGT, para Cabré a Terminologia deve partir da observação de textos reais, organizados num *corpus*. Nesses textos serão identificadas unidades de conhecimento especializado – morfemas, lexemas, sintagmas, orações e os próprios textos – de onde se extraem as “unidades terminológicas”, respeitando-se as necessidades específicas dos usuários a que se destina a aplicação. Essas unidades são então descritas em seus aspectos morfológicos, semânticos, sintagmáticos, funcionais e discursivo-pragmáticos. A partir da descrição é possível extrair generalizações, que são a base para a formulação de constantes, normas. O passo seguinte é elaborar os materiais, que podem atender a diversos segmentos da comunidade especializada: prática e ensino de tradução, jornalismo científico, ensino de língua para fins específicos, lexicografia, documentação, etc.

⁵⁴ por preocupar-se com aspectos semânticos da unidade terminológica

Para Cabré, o conteúdo de um texto especializado é como um mapa conceptual de um território (Cabré, 2005) – a informação estrutura-se em Unidades de Significação Especializadas que, quando representam um nó conceptual na estrutura de conhecimento de uma dada área especializada, são chamadas de Unidades Terminológicas (UT). Para a TCT, essas unidades são basicamente nominais, de carácter denominativo e valor referencial. Estas podem ser compostas de uma unidade lexical apenas (monoléticas / simples) ou mais de uma unidade (poliléticas / sintagmáticas).

O tratamento dado pela TCT à variação léxica (polissemia, sinonímia, homonímia, etc.) também é outro: os termos não mais *pertencem* a determinadas áreas, mas **são usados** em determinadas áreas. Seu sentido se constrói no texto, e daí a variação. Por isso, o percurso do trabalho terminológico, do ponto de vista do terminólogo, é semasiológico (do termo para o conceito), e não onomasiológico – como queria Wüster (ainda que o possa ser do ponto de vista do especialista).

Ao longo dos últimos 20 anos, a TCT foi sofrendo ajustes e evoluindo com as contribuições dadas por pesquisadores que aplicaram seus pressupostos na compilação de diferentes materiais terminológicos mono- e multilíngües, de forma que, hoje, dá conta de vários fenómenos do texto técnico de forma acurada e aprofundada. Vide, por exemplo, os trabalhos de excelente qualidade que têm sido produzidos pelo grupo de pesquisadores do IULA ou por alunos de mestrado e doutorado que por lá desenvolveram suas pesquisas.

Apesar da importância dada ao texto técnico por essas três correntes teóricas (TST, Socioterminologia e TCT), a perspectiva teórica que parece mais afinada com esta pesquisa é a Terminologia Textual, conforme procurei demonstrar no item a seguir.

1.3.3 Terminologia Textual

A Terminologia de orientação lingüístico-textual, ou “Terminologia Textual” (Bourigault e Slodzian, 2004 [1999]; Finatto, 2007: 222) considera como unidade de análise o texto técnico – e, portanto, as unidades maiores que os termos e os aspectos relacionados à comunicação especializada como um todo. É a linha teórica que mais se aproxima desta pesquisa pois tem na Lingüística de Corpus, atualmente, um de seus aportes metodológicos mais importantes.

Os fundamentos da Terminologia Textual baseiam-se numa “lingüística do texto especializado”, conforme proposta inicialmente por Hoffmann (vide Hoffman, 2004 [1988]) e desenvolvida em trabalhos posteriores seus e de autores como Kocourek, Bourigault, Slodizian, Condamines, L’Homme e Finatto, para citar apenas alguns. É a única teoria que parece permitir uma afirmação como a que estou propondo neste trabalho, qual seja, de que a pertinência ou não das unidades textuais a uma estrutura de conhecimento especializado, ou mesmo seu estatuto morfossintático-semântico, não é a pedra de toque que vai determinar se essas unidades merecem ou não ser estudadas. O fato de terem uma **presença significativa no texto técnico** (atestado empiricamente num *corpus* de estudo criteriosamente compilado) poderia, sim, ser considerado condição suficiente para que sejam objeto de análise e descrição.

(...) para além do reconhecimento da linguagem, aberto tal **reconhecimento a quaisquer itens do léxico** e ao reconhecimento de uma enunciação científica, sejam esses itens advérbios ou conectivos, qualquer glossário ou outro produto terminográfico, seja impresso ou sob forma de mecanismo interativo em um *website*, não pode afastar-se de uma postura descritiva. (Finatto, 2007: 229, grifo meu)

Para o tradutor, por exemplo, todos os recursos lingüísticos empregados para garantir a compreensão entre os profissionais de uma certa área de especialidade são de seu interesse, já que essas unidades são uma parte considerável do potencial expressivo de que dispõe para a correta tradução do conteúdo especializado engendrado por num texto técnico na língua de partida. E, como bem observaram Bourigault e Slodizian, há boas razões para pautar o trabalho terminológico na análise de textos da área:

A atividade de construção de uma terminologia torna-se, antes de mais nada, uma tarefa de análise de *corpora* textuais. Há duas razões fundamentais para isso:

- As aplicações da Terminologia são mais freqüentemente aplicações textuais (tradução, indexação, redação assistida); **a terminologia deve “vir” dos textos para melhor “retornar” a eles**. É justamente porque ela nunca é desvinculada do texto que se fala em “terminologia textual”.
- é nos textos produzidos ou utilizados por uma comunidade de especialistas que estão expressos e, portanto, disponíveis uma boa parte dos conhecimentos compartilhados por essa comunidade; é, pois, por aí que se

deve começar uma análise. (Bourigault e Slodzian, 2004 [1999]: 103, grifo meu)

Daí depreende-se a importância da Lingüística de Corpus para a Terminologia Textual – é ela que vai possibilitar a renovação teórica e metodológica de que a Terminologia necessita para poder dar conta da “massa dos dados analisados e dos prazos impostos” (Bourigault e Slodzian, 2004 [1999]: 104):

A atualidade da questão terminológica ao longo das transformações ocorridas, no que diz respeito à escala e ao ritmo de produção, bem como a dimensão das necessidades, exigem uma renovação teórica e metodológica. Permitindo abordar sistematicamente o estudo das práticas textuais reais, a lingüística de *corpus*, com suas técnicas e ferramentas, possibilita o acesso às expressões lingüísticas concretas de onde será possível fazer emergir e depois normatizar os termos pertinentes. (Bourigault e Slodzian, 2004 [1999]: 107)

Outro ponto importante levantado por esses autores – e também observado por Temmerman (2004 [2001]) – é a importância de se considerar o público-alvo como fator determinante em todas as etapas do trabalho terminográfico. Para eles, há tantas “terminologias” quantas forem suas aplicações.

(...) a constatação da variabilidade das terminologias se impõe: considerando uma área de atividade, não há UMA terminologia que represente O conhecimento dessa área, mas tantas terminologias quantas forem as aplicações nas quais estas forem utilizadas. Diferem quanto às unidades coletadas e quanto à descrição, conforme a aplicação visada. (Bourigault e Slodzian, 2004 [1999]: 103)

Os autores vão além, tocando no ponto fulcral que filia minha visão de tradução (ancorada no texto) a esse posicionamento teórico, ao sugerirem que, ao contrário do que postulam as teorias sobre a Terminologia, não existe um conhecimento especializado verdadeiro, único, ideal, que todo trabalho terminológico almejaria, em tese, acessar e descrever:

Os termos não são “unidades de conhecimento” que viriam “habitar a língua”. A tarefa de análise terminológica não é, pois, um exercício de redescoberta de um sistema nocional preexistente que caracterizaria a área. As noções não têm anterioridade ou

prioridade em relação às palavras: a terminologização é um processo paralelo à elaboração conceitual. (Bourigault e Slodzian, 2004 [1999]: 106)

Contrária a uma abordagem estreitamente onomasiológica, “a pedra no sapato” do lingüista, a abordagem textual abre as portas a todas as experiências da análise lingüística textual (ultrapassando-se assim a visão limitada das LSPs [Linguagens para fins específicos]). (*op. cit.*: 107)

Auger e L’Homme (2004 [1994-5]: 110) corroboram Bourigault e Slodzian no tocante às limitações da terminologia tradicional para lidar com as relações entre termo e texto especializado: “(...) os aspectos que tangem aos outros aspectos lexicais (vocabulário além do terminológico), o funcionamento do termo no texto especializado e suas relações com o texto foram negligenciados”. No enfoque adotado pelas teorias terminológicas conceptualistas, para as quais a tarefa terminológica é puramente onomasiológica (parte do conceito para o termo), não há lugar para relações não-hierárquicas ou puramente textuais; por essa razão, o uso de *corpora* como ponto de partida é incompatível com essa visão de Terminologia:

Terminology structuring in conceptual approaches is often carried out in order to **represent knowledge and not linguistic units**. Problems arise when this work is done using corpora as a starting point, since linguistic units (such as terms) do not behave in a way that reflects perfectly a given knowledge structure. When analyzing terms, considerations regarding knowledge structure will constantly interfere with factors related to the behaviour of linguistic units in text.⁵⁵ (L’Homme, 2004 [1994-5], grifo meu).

Nesse mesmo sentido, mas dando um passo além – e na mesma direção que estou apontando na presente pesquisa –, Finatto (2007) sugere que a Lingüística de Corpus deve ser vista não somente como uma metodologia de trabalho adequada para a Terminologia, como sugeriram Bourigault e Slodzian, mas como abordagem teórica, já que subsume uma visão de língua harmonizada com a perspectiva comunicativo-textual da Terminologia Textual:

⁵⁵ A estruturação da terminologia nas abordagens conceptuais é levada a cabo, geralmente, com vistas a representar o conhecimento, e não as unidades lingüísticas. Problemas surgem quando esse trabalho é feito usando-se corpora como ponto de partida, já que unidades lingüísticas (como os termos) não se comportam de uma maneira que reflita perfeitamente uma estrutura do conhecimento. Na análise de termos, considerações acerca da estrutura do conhecimento interferirão constantemente com os fatores relacionados ao comportamento das unidades lingüísticas no texto.

(...) para haver uma boa aproximação entre a informatização e a Terminologia, via LC [Linguística de Corpus], uma aproximação que ultrapasse a visão da LC como mera metodologia, “contagem estatística de palavras” ou o uso de um *software*, é preciso que retomemos criticamente nossos princípios básicos. Afinal, a LC não é apenas um método, nela há uma concepção de língua como sistema probabilístico e a percepção de que a língua que mais interessa é a língua em uso e não a língua mente-cérebro, antes de ser verbalizada. (Finatto, 2007: 228)

A Linguística de Corpus não deve, portanto, ser utilizada simplesmente como um artifício para automatizar as tarefas tradicionais do trabalho terminográfico – ela deve mesmo redefinir o *que* deve ser compilado e, por conseguinte, as metodologias de coleta e registro dessas unidades que tipificam os textos técnicos.

A informatização, mediada pela LC [Linguística de Corpus], mostra para a Terminologia que a melhor observação da linguagem é, sim, extensiva. (...) Pela lente de aumento gerada pela observação extensiva, evidenciam-se fortes características do texto científico colocadas por alguns elementos na superfície gramatical, mas que **só serão percebidos se formos capazes de extrapolar a fronteira do que seja “estritamente terminológico”**. (Finatto, 2007: 226-7, grifo meu)

(...) terms behave like other lexical units and must be dealt with accordingly. Terms will acquire their specificity through a given application with set objectives, but **as units occurring in corpora, terms cannot be differentiated from other lexical units**⁵⁶. (L’Homme, 2004, grifo meu)

Finatto, no artigo já citado de 2007 (p. 223), dialoga com Cabré ao remotivar a metáfora proposta pela autora na Teoria das Portas, ampliando-a. Para Finatto, a linguagem especializada pode ser comparada a um edifício, repleto de divisões: salas, andares, etc. E a Linguística de Corpus seria a chave para abrir uma nova porta de acesso a esse edifício.

Por me interessar por tudo aquilo que não necessariamente é “estritamente terminológico” no texto técnico, e por acreditar que o texto técnico deve ser, tratado acima de tudo, como texto – e não como repositório de conceitos especializados – em vez de adotar metodologias disponíveis na terminografia atual para a compilação de *corpora* e coleta das unidades a comporem a amostra de dicionário voltado para o

⁵⁶ (...) os termos se comportam como qualquer outra unidade lexical, e devem ser tratados como tal. Eles vão adquirir sua especificidade de acordo com uma dada aplicação e seus objetivos preestabelecidos, mas na condição de unidades ocorrendo em corpora, os termos não se distinguem de outras unidades léxicas.

tradutor, proponho utilizar a Lingüística de Corpus como abordagem teórica e metodológica para executar essas etapas, conforme discutirei no Capítulo III.

Assim, pode-se dizer que, a presente pesquisa, ainda que se utilize da Lingüística de Corpus, ela não servirá de “chave”, pois não me interessa aqui adentrar a “casa” do termo, conhecê-lo internamente – essa tarefa será reservada à Terminologia, que a tem realizado, com maior ou menor êxito, de diferentes perspectivas teóricas, como as que acabei de destacar. A Lingüística de Corpus me servirá, sim, de lupa (ou microscópio, para usar uma metáfora corrente entre os pesquisadores desta área), para que a tessitura acidentada desses vários textos-mapa (para usar outra analogia de Cabré) que compõem o atlas vivo da comunicação especializada representado pelo corpus, possa ser examinada em detalhe. Nessa análise, me ocuparei não apenas da descrição de “casas” e “edifícios”, mas de tudo o que saltar à vista na paisagem textual mutável e acidentada em que figuram as palavras.

A proposta do uso da Lingüística de Corpus como abordagem na compilação de dicionários voltados para tradutores será apresentada nos Capítulos III e IV. No capítulo que se segue, apresento uma caracterização das tipologias textuais do gênero Culinária para, em seguida, discutir as necessidades específicas do tradutor técnico que trabalha com a Tradução Culinária no Brasil e as lacunas dos materiais de consulta disponíveis na área no mercado editorial brasileiro atualmente.

This document was created with Win2PDF available at <http://www.daneprairie.com>.
The unregistered version of Win2PDF is for evaluation or non-commercial use only.

Capítulo II

Tradução culinária no Brasil: mercado de trabalho, materiais de apoio e suas lacunas

Para falar do papel do dicionário técnico na tradução culinária e das carências dos materiais de consulta disponíveis no Brasil, falarei antes sobre o que caracteriza uma receita, que implicações estas características têm para o fazer tradutório, e que status têm o tradutor e as traduções da área no mercado editorial e de tradução atualmente. Uma vez que esse status está intimamente ligado ao valor que a sociedade dá ao tema em si, inicio o capítulo com um breve apanhado dos principais fatos que colocaram a Culinária na posição de destaque que ocupa hoje no cenário brasileiro e internacional.

2.1 Globalização e o *boom* da Culinária

Comida “globalizada” não é uma exclusividade dos tempos modernos. Os alimentos e preparações sempre acompanharam o ser humano em suas viagens, êxodos e conquistas, desde tempos imemoriais, como atestam diversos historiadores. No entanto, há certos momentos históricos em que a humanidade presenciou uma interação mais forte entre os povos e, conseqüentemente, um intercâmbio maior de conhecimentos e ingredientes culinários.

A expansão do budismo na Ásia a partir do século I, e do Islam do oeste mediterrâneo para a Índia entre os anos 650 e 850, o Império Romano, a criação do Império Otomano em 1300, ou mesmo as revoluções Francesa e Americana do final do século XVIII podem ser citados como exemplos de “globalização”, ou seja, momentos em que houve internacionalização, deslocamentos de povos e intercâmbios de costumes. Mas foi o período das grandes navegações – ou “primeira globalização” (Carneiro, 2004: 153) –, a partir do século XV, que difundiu em larga escala os chamados *alimentos modernos*: açúcar, especiarias, bebidas quentes, batata, tomate, milho, arroz, trigo, etc., que constituem a base da alimentação atual na maior parte do mundo.

A globalização *de facto*, bem mais recente, é um fenômeno essencialmente econômico, cujos primórdios remontam à Revolução Industrial, quando o sistema de produção capitalista, a divisão de trabalho e a setorização da economia começaram a se estabelecer. Fortaleceu-se após a Segunda Guerra Mundial, quando foram criados vários órgãos e acordos internacionais para promover o livre comércio entre as nações, favorecendo o desenvolvimento econômico – especialmente dos países e instituições financeiras e políticas detentoras do poder.

Os enormes progressos obtidos nos setores de tecnologia e meios de comunicação, especialmente a popularização do computador, a partir do início da década de 1980, e o posterior surgimento da Internet, consolidaram de vez a sociedade global, alterando todas as esferas da vida humana. Gomes e Barbosa assim descrevem os principais marcos desse movimento em direção à globalização, no Brasil, no que respeita o comportamento alimentar:

(...) em 1950 se deu a implantação da “indústria branca” (fogões, geladeiras, liquidificadores, batedeiras etc.), cujo ponto culminante foi a década de 1970, com o *boom* dos alimentos industrializados

no Brasil, incluindo a introdução e a relativa difusão do *fast food* (Gomes e Barbosa, 2004).

Atualmente existem cerca de 1,5 milhão de estabelecimentos que comercializam alimentos e bebidas diretamente ao consumidor no Brasil – um mercado que cresceu 229,8% em apenas dez anos. A projeção deste nicho para 2008 é de 6,6% de crescimento, com um faturamento superior a R\$ 100 bilhões⁵⁷.

Vários fatores colaboraram para a ascensão da temática Culinária junto à sociedade brasileira a partir do início dos anos de 1990, entre eles: a grande oferta de informação proporcionada pela Internet; a criação de programas televisivos sobre o tema para a programação da TV aberta e a cabo (muitos deles comandados por *chefs* estrangeiros radicados no Brasil, como Claude Troisgros, Emmanuel Bassoleil e Olivier Anquier); o aumento do poder aquisitivo e da qualidade de vida da população (especialmente das classes média e média-alta) devido ao controle da inflação, o que impulsionou o setor de alimentos como um todo; o aumento das importações do setor, especialmente a partir do governo Collor, quando se intensificou no país a abertura comercial a produtos estrangeiros; o fortalecimento do turismo de massa e, conseqüentemente, do turismo gastronômico; a oferta de cursos de formação profissional, que inseriram a Culinária no cenário do ensino superior.

Seguindo as tendências internacionais, a Culinária – chamada agora de Gastronomia⁵⁸, em respeito a seu novo status social e intelectual – foi, aos poucos, ganhando maior visibilidade junto à elite intelectual e acadêmica do Brasil, que passou discutir seus temas em congressos e eventos organizados pelo país afora. Em 2004, realizou-se, em Brasília, o 1o. Encontro Brasileiro de Gastronomia e Segurança Alimentar, com apoio da UnB (Universidade de Brasília) e do Centro de Excelência em Turismo, que reuniu personalidades do cenário gastronômico brasileiro e internacional. Os assuntos debatidos, conforme anuncia a Apresentação da *Coletânea de Palestras* publicada posteriormente (Araújo *et al.*, 2004), vão “muito além da combinação de técnicas culinárias e degustação”. Os 19 artigos dividem-se entre os temas: *hábitos alimentares, cenário nacional da gastronomia, gastronomia e globalização, segurança*

⁵⁷ <http://www.portocultura.com.br/gastronomia/?id=7&idNot=1650>

⁵⁸ A meu ver, o termo Gastronomia é hipônimo do termo Culinária, isto é, uma particularização desse fenômeno mais abrangente, envolvendo a alimentação humana como um todo, que estou chamando de Culinária: mais especificamente, é uma versão institucionalizada da culinária. No entanto, no presente trabalho, ora os dois termos serão usados um pelo outro, ora para diferenciar o aspecto mais prático desse fenômeno (=Culinária) do mais intelectualizado, de maior prestígio social (= Gastronomia).

alimentar - acesso à alimentação, tendências na gastronomia e garantia da qualidade na gastronomia. Mais um fato positivo sobre a iniciativa: o livro, de 317 páginas, pode ser consultado gratuitamente na Internet (vide bibliografia).

Em 2006, os chefs Alex Atala, Edinho Engel e Mara Salles reuniram-se por três dias para realizar o 1o. Laboratório de Cozinha Brasileira (com uma segunda edição em 2008). De acordo com a reportagem publicada pelo suplemento *Paladar* da Folha de São Paulo, dedicado exclusivamente à Gastronomia, esses profissionais propõem uma “Declaração de Independência” para a cozinha brasileira, com iniciativas que vão desde a definição e o estudo dos “biomas” culinários nacionais (regiões demarcadas de acordo com os ingredientes mais usados em suas preparações, e não mais segundo seu posicionamento geográfico), passando pela difusão dos ingredientes e pratos brasileiros no cenário internacional, a sistematização de receitas e procedimentos típicos, a promoção de encontros entre *chefs*, pensadores, cozinheiros, quituteiros, autodidatas e demais profissionais brasileiros envolvidos no setor, entre outros.

Eventos como esses exemplificam a importância adquirida pela Culinária junto à sociedade brasileira. Prova ainda mais óbvia é o espaço que o tema passou a ocupar em todos os tipos de mídia nos últimos 10 ou 15 anos: programas de TV; livros; suplementos em jornais de grande circulação, fartamente ilustrados; revistas especializadas (ou não – a edição de jul/ago/set 2002 da revista trimestral que o plano de Saúde Unimed envia para seus clientes via correio trouxe, na seção que fecha a publicação – de nome “Aquarela” –, a matéria “Dicas de gastronomia para os dias frios”). Até mesmo a edição do boletim eletrônico do Banco Citibank “Dirigido aos Clientes CitiGold”, de abril de 2002, publicou, em meio aos tópicos “Rentabilidade, Macroanálise, Investimentos e Seguros”, o item “Gastronomia”, com a receita de um *Risotto alla Milanese*.

Em setembro de 2004, o jornal *O Estado de São Paulo*, em matéria de capa do suplemento de domingo “Telejornal”, publicou a matéria “Como ganhar a audiência pelo estômago”, em que afirma:

Nada tem mais presença na TV do que a gastronomia, tema que vem invadindo também a ficção, com destaque cada vez maior. Em ‘Senhora do Destino’ [novela exibida pela Rede Globo no horário nobre], há personagens e cenários exclusivamente dedicados à arte de cozinhar, o que renderá o lançamento de um livro de receitas (O Estado de S. Paulo, seção Telejornal, 12/09/2004)

O *boom* da Culinária na imprensa brasileira chama a atenção também de pesquisadores da área de Turismo e Divulgação:

(...) a gastronomia vem conquistando espaço no jornalismo e na televisão. Proliferam programas dedicados a oferecer uma hora na cozinha com *chefs* de todo o mundo. (...) A comida continua a ser atrativo principal dos serviços oferecidos pelo *trade* de turismo. (Montoro, 2004: 114).

Em seu artigo, Tânia Montoro analisa os cadernos de Turismo “Lugares” (Correio Braziliense), “Boa Viagem” (O Globo) e “Folha Turismo” (Folha de São Paulo) de março a abril de 2004 e constata que o espaço dedicado à Gastronomia nessas mídias no período foi de cerca de 30%. Quanto ao teor, as matérias versaram sobre “conteúdos promocionais de restaurantes, alimentos ou bebidas, excelência em serviços da hotelaria e destinos turísticos movidos pelas particularidades / especificidades da gastronomia” (Montoro, 2004: 97).

A autora lamenta a “superficialidade noticiosa” dos conteúdos veiculados que, além de promoverem a “mercantilização da cultura e da natureza”, não dão voz aos inúmeros profissionais envolvidos na gastronomia: “Químicos, nutricionistas, agentes sanitários, agricultores, tecnólogos de alimentos, biólogos, bioquímicos, médicos não se configuram como atores e fontes para os jornalistas e editores das matérias sobre gastronomia” (*op. cit.*: 113).

Essa percepção de que a Culinária não se restringe ao preparo e ao consumo de alimentos parece não ser ainda ponto pacífico no imaginário e no intelecto do brasileiro. Ainda que a maioria dos jornais atuais conte com um suplemento voltado ao tema (ou dediquem um espaço considerável a esse assunto em seus cadernos de comportamento, entretenimento, turismo, etc.), há quem considere o assunto como corriqueiro, de menor monta, restrito aos grupos profissionais do setor alimentício e à esfera do lar. Um dos motivos para isso é o fato de ser um fenômeno muito mais recente no Brasil do que em outros países.

Nos Estados Unidos, por exemplo, o *boom* da Culinária ocorreu em proporções muito maiores, e alguns anos antes de eclodir no Brasil. O país tem um canal totalmente dedicado à culinária desde 1993, o *Food TV* (e vários outros cuja programação inclui programas com temas relacionados à culinária), um sem número de revistas especializadas (*Bon Appétit*, *Cook’s Illustrated*, *Food & Wine*, *Gonomica*,

Gourmet, *Martha Stewart Living*, *Slow Food*, etc.), redes de lojas que vendem especialidades culinárias, inclusive pela *Internet* (Dean & DeLuca, Williams-Sonoma, Whole Foods, etc.), editoras voltadas apenas à publicação de livros de Culinária (Favorite Recipes Press, Wimmer Cookbook Distribution, Cookbook Publishers, Morris Press Cookbooks, etc.). A tradicional escola de gastronomia francesa Le Cordon Bleu conta atualmente com 14 franquias espalhadas pelos principais centros urbanos dos Estados Unidos.

If there's a major difference between now and the sixties and seventies, it's that the scale is so much larger; culinary sophistication is no longer the province of a tiny gourmet elite. The historically unrivaled run of prosperity in the United States in the eighties and nineties (...) has led to the creation of an expanded leisure class that treats food as a cultural pastime, something you can follow the way you follow sports or movies.⁵⁹
(Kamp, 2006: xiii)

A presença do tema é tão massiva no dia-a-dia dos americanos que tem ditado tendências na sociedade, principalmente nas classes mais abastadas (já que as menos favorecidas vêm sofrendo, concomitantemente, com um problema de obesidade de proporções epidêmicas). O livro escrito pela *personal chef* da celebridade Ophra, de acordo com a Ipsos-NPD (empresa americana que pesquisa exclusivamente o comportamento do mercado editorial do segmento gastronômico), foi o livro de culinária que vendeu mais cópias até hoje nos Estados Unidos – 5,4 milhões de exemplares (Jacob, 2005: 235).

You can be a non-cook and still be a food obsessive, attending new restaurant openings like a theatergoer (...). Or you can be a serious cook and fill your kitchen with professional-grade equipment – the six-burner Viking Range, the Sub-Zero refrigerator, the supersharp Global knives from Japan – and buy your produce and artisanal edibles from the local farmers' market that didn't exist ten years ago, and exhibit your disdain for mere "feeders" by subscribing to and following the recipes of *Cook's Illustrated*, the meticulously researched, trend-averse

⁵⁹ Se há uma diferença significativa entre o momento atual e os anos sessenta e setenta, ela está na proporção, muito maior agora; a sofisticação culinária não é o domínio apenas de uma elite de meia dúzia de gourmets. A prosperidade sem precedente histórico dos Estados Unidos durante os anos de oitenta e noventa (...) propiciou a criação de uma categoria recreativa cujos numerosos seguidores tratam a comida como um passatempo cultural, algo que você pode acompanhar como quem acompanha os esportes ou o cinema.

anti-glossy that thrived since its founding in 1980 without the benefit of any advertising revenue.⁶⁰ (Kamp, 2006: xiii)

Com o interesse crescente pela Culinária e pela Gastronomia em todo o mundo, o mercado editorial vinculado ao setor se expandiu:

(...) it can be said that cookbooks have been the ugly duckling of publishing. Ten years ago, cookbooks were one of the least respected sectors by publishers, very rarely mentioned at all. (...) Today, with the cookbook boom, publishers take cookbooks very seriously.⁶¹ (Gourmand World Cookbook Award, 2005 edition)

Isso afetou diretamente o mercado de tradução culinária, como procuro mostrar nos subitens a seguir.

2.1.1 A Culinária no mercado editorial

De acordo com dados da *Gourmand World Cookbook Award*, uma entidade dedicada a selecionar os melhores livros de Culinária publicados no mundo anualmente, em 2007 foram editados 26.000 títulos sobre o tema no mundo todo (24.000 em 2005). Segundo o fundador e presidente dessa entidade, Edouard Cointreau, vários fatores colaboraram para o *boom* da culinária no mercado editorial internacional:

Sales of cookbooks in the world increase at a steady 5% a year since 1994. Production of new titles is 24.000 a year, the double of 10 years ago. The reasons for this boom are:

- Food television success, led by BBC (UK) and PBS (US). There is now at least one food cable network in each country, such as the TV Food Network (US), Canal Cocina (Spain), Gourmet TV or Cuisine TV in France, etc.

⁶⁰ Nos Estados Unidos, você pode ser um não-cozinheiro e ainda ser obcecado por comida, comparecendo assiduamente a cada inauguração de restaurante como quem vai ao teatro (...). Ou, você pode ser um cozinheiro de verdade e encher sua cozinha com equipamentos profissionais – um fogão Viking de 6 bocas, uma geladeira Sub-Zero, as afiadíssimas facas Global do Japão – e comprar suas frutas, legumes, verduras e ingredientes artesanais numa feira de produtores locais, que não existiam há dez anos, e demonstrar seu desdém pelas pessoas que “comem para viver” assinando e seguindo as receitas da *Cook’s Illustrated*, revista de conteúdo meticulosamente pesquisado, avessa aos modismos e ao papel fotográfico brilhante, que tem sobrevivido desde sua publicação, em 1980, sem o aporte de qualquer tipo de material publicitário.

⁶¹ (...) pode-se dizer que os livros de culinária foram, por muito tempo, o patinho feio do mercado editorial. Há dez anos, constituíam um dos segmentos mais desrespeitados pelos editores, raramente mencionado, quando muito. (...) Atualmente, com o boom dos livros de culinária, editores estão levando essas publicações muito a sério.

- Nutrition and food health awareness have risen everywhere. Asia always made it very important, it is now a key factor in most northern countries.
- Food has a strong comforting role in a world increasingly perceived as dangerous.
- New book production technology make cookbooks photography, colour and good paper increasingly easier, quicker and cheaper. (Gourmand World Cookbook Award, 2005)⁶²

Em artigo escrito em 2003, as antropólogas Laura Graziela Gomes e Livia Barbosa fazem um estudo, no Brasil, do que denominam *a culinária de papel*, i. e.: “todo material impresso sobre a arte de cozinhar e/ou relativo às diferentes cozinhas e culinárias existentes” (Gomes e Barbosa, 2004: 1). Seu objetivo é determinar diacronicamente como a “culinária de papel” influenciou o processo de formação de uma cultura gastronômica no país e por ele foi influenciada. E fazem isso por meio da análise dos títulos de todas as obras publicadas em solo nacional sobre Culinária e Gastronomia, do início do século XX ao início do século XXI. O levantamento é feito junto à Biblioteca Nacional e a sebos e livrarias de Belo Horizonte e do Rio de Janeiro, chegando a um total de 907 obras, entre nacionais e traduzidas.

As autoras relatam que 493 obras foram escritas por mulheres, 237 por homens (menos de 20 títulos até 1970), 30 são de autoria coletiva e 65 de autoria desconhecida ou institucional. Somente em cerca de 50% delas (as mais recentes) a data de publicação pode ser claramente identificada. Com base nas observações das autoras quanto às tendências históricas da Culinária e da Gastronomia no Brasil, de acordo com os títulos de livros analisados, construí um quadro-resumo simplificado:

⁶² A venda de livros de culinária no mundo teve um crescimento estável de 5% ao ano desde 1994. O lançamento de novos títulos gira em torno de 24.000 ao ano, o dobro de obras publicadas há 10 anos. As razões para esse *boom* são: - O sucesso dos canais de culinária, encabeçado pela BBC britânica e pelo canal PBS norte-americano. Atualmente, há pelo menos um canal de culinária na TV a cabo de países como os Estados Unidos (Food Network), a Espanha (Canal Cocina), a França (Gourmet TV ou Cuisine TV), etc.; - A conscientização sobre a nutrição e a alimentação correta intensificou-se de forma geral. A Ásia sempre valorizou esses aspectos, mas agora é um fator chave na maioria dos países do hemisfério norte.

décad as	1930-1940	1950-1960	1970	1980	1990	2000
equipamentos	métodos artesanais	fogão, geladeira, panela de pressão, liquidificador, batedeira		freezer	processadores, forno microondas, utensílios especiais	
temáticas	doce/salgado, festa/dia-a-dia, cru/cozido, cozido(panela)/assado(forno)			especialização: livros dedicados a um só alimento, ou tipo específico de preparação		
termos preferidos	“cozinha / culinária” (= técnica / espaço físico / tipo de preparação / pratos regionais)				“gastronomia” (= cozinha internacional, universalista e cosmopolita; arte)	
teor das obras	cozinha completa, segredos culinários	cozinha rápida, prática, econômica; técnicas de preparo (utilizando novos equipamentos)			cozinha <i>light</i> , <i>diet</i> , internacional	cozinha <i>fusion</i> , macrobiótica, nacional
foco	lar (vínculos afetivos)		casa (equipamentos/praticidade)		indivíduo (pazer/hobby)	
público-alvo	cozinheira, dona-de-casa, “rainha do lar” (lar = família, vínculos afetivos)			mulher (cozinha prática / equipada)	mulher (corpo/dieta), homem (lazer/hobby)	
cozinha	lugar pouco nobre, reservado aos empregados			ambiente agradável, nobre, denota status socio- econômico do dono. palco de reuniões sociais		

Figura 3: quadro-resumo baseado em Gomes e Barbosa (2004)

Em vista dos dados numéricos obtidos nesse levantamento, as autoras consideraram, à época (2004), que “pouco ou quase nada se escreveu sobre esse tipo de literatura” no país, ainda que tenham observado um aumento gradativo no número de publicações ao longo do século, que cresce consideravelmente após a década de 1990:

No período que vai do início da década de 1990 até 2003, registramos, até o momento, um total de 309 novos livros de culinária, sem considerar os títulos em língua estrangeira postos à disposição do público brasileiro. Esse número equivale a mais que o dobro de tudo aquilo que registramos como disponível do início do século XX até o final da década de 1980. (Gomes e Barbosa, 2004: 22)

Carlos Alberto Dória confirma essa tendência no artigo “A antropologia do mingau”, de 2006. Em pesquisa feita na Plataforma Lattes, “fonte web por excelência da produção acadêmica brasileira” (Dória, 2006: 207), o autor identificou cerca de 40 trabalhos acadêmicos que se debruçam sobre os temas Culinária e Gastronomia de diferentes perspectivas: antropológica, histórica, lingüística, nutricional, étnica, etc. Nessas iniciativas o autor vê um “enorme salto de qualidade que, silenciosamente, vem se processando”. E comenta, a esse respeito:

Sem dúvida são muitas as barreiras que se encontram em fase de superação. A primeira e principal delas é ideológica: a atitude tradicional das classes médias em relação ao que se passa dentro da cozinha. Como herdeira de preconceitos de origem colonial, para ela “cozinhar” sempre foi um verbo considerado menor, conjugado por gente que lida com as mãos e não com o intelecto. Só quando a formação de chef se tornou uma “carreira” para os filhos das classes médias é que a universidade pôde dedicar maior atenção à culinária. Mesmo assim, a disciplina ainda não possui enquadramento claro, e no seu sem-lugar “contamina” os vários nichos especializados antes referidos. (Dória, 2006: 208)

Essa atitude a que Dória se refere pode ser percebida subliminarmente em vários artigos de divulgação disponíveis na Internet. Na matéria “Estante culinária”, publicada no suplemento *Comer & Beber*, dedicado somente à Culinária, do jornal *Diário do Nordeste*, embora esteja justamente defendendo a interdisciplinaridade de culinária, o autor usa a expressão “metem a colher no assunto” para descrever o interesse de “Literatos e acadêmicos” pelo tema:

Engana-se, porém, quem imagina que a gastronomia é tema restrito a *chefs* e *gourmets*. Literatos e acadêmicos, eventualmente, metem a colher no assunto. Afinal, a cultura alimentar e os ritos à mesa há tempos mobilizam o interesse de disciplinas como a História, a Sociologia e a Antropologia, dentre outras. (Ricardo, 2007)

Irônico ou não, o uso da expressão “meter a colher” tem uma prosódia semântica extremamente negativa, pois o que vem à mente, imediatamente, é a negação “**não** se mete a colher...”. Ainda assim, o mesmo artigo dá mostra da miríade de livros publicados (e reeditados) nos últimos anos sobre o assunto por importantes editoras do país. Os mais acadêmicos reúnem diversas perspectivas teóricas (por exemplo, jornalismo culinário, crítica gastronômica, historiografia da ciência, turismo e hotelaria, sociologia). Há também inúmeros livros de receitas, muitos voltados à culinária brasileira e regional (vários deles bilíngües), além de reedições e traduções de obras históricas (como Câmara Cascudo e Alexandre Dumas).

Segundo pesquisa da Câmara Brasileira do Livro, em parceria com o Sindicato Nacional dos Editores de Livros, os títulos sobre gastronomia registram um crescimento de mais de 23% somente em 2006 no Brasil. Uma pesquisa feita em 10/08/2008 no *site*

da Livraria Saraiva⁶³ revelou um total de 4.584 títulos sob a categoria “culinária”, dos quais 1.584 são editados no Brasil e 3.000 importados. Boa parte dos títulos “importados” é, na verdade, editada e publicada em Portugal e vendida no Brasil, muitas vezes sem indicação dessa característica peculiar – um problema para o leitor desavisado (Teixeira, 2004). A Livraria Cultura possui um catálogo com mais de 2.000 títulos sobre o tema, dos quais cerca de 50% são importados⁶⁴.

Esses livros importados, ainda que estejam ocupando o espaço de produções nacionais, parecem já não mais intimidar as editoras do país, que passaram a publicar muito mais na área nos últimos anos. Somente a Editora Senac⁶⁵ conta com cerca de 140 títulos relacionados à Culinária / Gastronomia atualmente, muitos deles edições bilíngües, e a grande maioria de autoria de escritores e profissionais brasileiros.

Em 2002 foi inaugurada em São Paulo a primeira livraria especializada em livros de Culinária da América Latina, a Mille Foglie. Com cerca de 4.000 títulos em seu acervo inicial, oferecia desde livros de receitas para cachorros a obras com valor superior a R\$ 1.500,00 – a maioria em inglês. A empresária e proprietária Gabriela Mascioli explicou à Revista NutriNews⁹ o que a motivou a investir num negócio desse tipo no Brasil:

Percebi que o segmento da gastronomia cresceu muito nos últimos anos, constatee o aumento do número de restaurantes e conseqüentemente de *chefs* e escolas. Por outro lado vi vários lançamentos de equipamentos para cozinhas *gourmets*, o que significava que muitas pessoas estavam aderindo ao fenômeno *nesting*, muito usual no exterior, ou seja, muita gente começou a receber amigos em casa, e nada como receber com uma boa comida.⁶⁶

Voltada para as classes média e média-alta⁶⁷, a livraria, que começou vendendo cerca de 400 livros por mês e tinha nos *chefs* de todo o país seu público-alvo mais assíduo, fechou as portas em 2007. Apesar desse aparente fracasso, a seção de Culinária continua a ser a “vitrine” de todas as *megastores* (Fnac, Cultura, Saraiva, Nobel, etc.), livrarias e sebos do país, pois existe uma demanda crescente por essas

⁶³ <http://www.livrariasaraiva.com.br/>

⁶⁴ <http://www.diarioon.com.br/arquivo/3572/cadernos/viver-5134.htm> [informação sem data - provavelmente de 2003; acesso em: 18/8/2008]

⁶⁵ <http://www.editorasenacsp.com.br/>

⁶⁶ <http://www.nutrinews.com.br/edicoes/Mat06Ed195MilleFoglie.html> [acesso em 18/8/2008]

⁶⁷ “‘Tem cliente que gasta R\$ 15 mil de uma vez só’, diz Marcela Herz, vendedora da Mille Foglie (...)” (<http://revistaepoca.globo.com/Revista/Epoca/0,,EDR69071-5990,00.html>, acesso em 18/8/2008)

obras também junto à população de poder aquisitivo mais baixo. Mas, para que tenham apelo para esse público maior, as obras, quando importadas, devem ser traduzidas – o que já foi percebido por muitos investidores do setor no exterior e no Brasil:

Cookbook translations are increasing everywhere. El Bulli books or Ducasse in English at the top end of the market, over 150 Euros, are selling by the thousands: this was thought to be impossible five years ago when the top price was 50 Dollars. There are more quality French pastry books published in Japanese than in French. Half of the 400 cookbooks published annually in Sweden are translations. (...) Food and wine corporations as well as other cookbook sponsors now need to access international markets, and publish at the same time in at least 2 languages, always including English: successful publishers have to have access to several markets.⁶⁸ (Gourmand World Cookbook Award, 2005)

Uma enxurrada de livros de culinária vem tomando espaço nas vitrines de livrarias, revelando uma curiosa tendência: livro de receitas vende bem. E mais: não vende só para mulheres. Há um público diversificado de olho nos lançamentos desse setor, desde as coletâneas sofisticadas até os manuais pragmáticos. Da dona de casa que se aventura por complicadas receitas, só para agradar a família, ao executivo descasado, que necessita de socorro para ligar um microondas, há títulos para todos os gostos e urgências. Comprar comida "impressa" é moda que pegou. (Barone, 1999).

Isso significa que há muito trabalho pela frente para o tradutor da área no par de línguas inglês-português, e não só nessa direção (o inglês é língua em que mais se publica sobre Culinária atualmente), mas também na direção português-inglês, haja vista o interesse crescente do público-alvo da área pelas culinárias regionais de todo o mundo. Mas será que as traduções Culinárias, a exemplo do que aconteceu nos mercados editoriais internacionais, tem recebido a mesma atenção e consideração por parte de tradutores, editoras e agências de tradução no Brasil?

⁶⁸ *A tradução de livros de culinária está aumentando em todo o mundo. Os livros do El Bulli ou os de Ducasse em inglês, no topo da lista e custando mais de 150 euros, estão vendendo aos milhares: imaginava-se que isso era impossível há cinco anos, quando o preço mais alto era 50 dólares. Há mais livros de qualidade sobre a confeitaria francesa publicados em japonês do que em francês. Metade dos 400 livros de culinária publicados anualmente na Suécia são traduções. (...) As corporações do setor de Alimentos e bebidas, bem como outros patrocinadores de livros da área, agora precisam entrar nos mercados externos, e publicar ao mesmo tempo em pelo menos 2 línguas, sempre incluindo o inglês: os editores bem-sucedidos têm de ter acesso a vários mercados.*

2.1.2 A Culinária no mercado de tradução brasileiro

A profissão de tradutor no Brasil, ainda que seja reconhecida pelo Ministério do Trabalho, não foi regulamentada até o momento, isto é, não existe um órgão oficial que aplique uma prova de regulamentação, como fazem a OAB, o CRM, etc. Isso significa que qualquer pessoa pode ser tradutor, mesmo sem ter tido uma educação formal para exercer essa profissão (ou até mesmo sem ter completado o ensino fundamental).

Os cursos de formação de tradutores no Brasil, na contra-corrente do momento histórico de expansão vivido pelo setor editorial no país, estão se tornando cada vez mais escassos – haja vista o recente fechamento do Curso de Especialização em Tradução de Longa Duração da USP, em nível de pós-graduação *lato sensu*, que por cerca de 25 anos formou profissionais e pesquisadores da área.

A despeito de todos esses problemas e limitações, o mercado de trabalho em tradução, especialmente em tradução técnica e nos pares de língua que envolvem o inglês (e mais recentemente, o espanhol), tem se alargado no país. E a tradução culinária, com a expansão vivida pelo setor de alimentos de modo geral, tem figurado, com frequência cada vez maior, entre as áreas especializadas do mercado de tradução.

O mercado de tradução na área de Culinária não se restringe aos livros de receitas. Além de traduzir receitas – que não estão só em livros culinários, mas em toda parte (*websites*, embalagens de produtos, todo tipo de mídia: jornal, televisão, rádio, revista, obras literárias, músicas e até mesmo cartões telefônicos⁶⁹ – vide item 2.2) – o profissional bem treinado pode atuar em vários setores da indústria de alimentos e bebidas; em empresas de importação e exportação de gêneros alimentícios; junto a estabelecimentos que servem alimentos e bebidas para turistas estrangeiros, ou associações de bares e restaurantes; em editoras, na coordenação de projetos de tradução da área; no mercado de audiovisual, fazendo a legendagem de programas de Culinária internacionais; junto a escolas de ensino técnico e superior da área, auxiliando na tradução e pesquisa do material didático, sob a coordenação do corpo docente, etc.

Nas próprias editoras, a demanda por profissionais especializados na área tem aumentado não apenas para dar conta do número crescente de livros de culinária

⁶⁹ esses cartões circularam no mercado em 2002: traziam em seu verso receitas de coquetéis e pratos típicos brasileiros cuja foto vinha estampada no anverso.

importados que têm sido publicados no Brasil, mas também para traduzir outros tipos de livros relacionados à área – memórias, romances, diários de viagens, guias turísticos, obras acadêmicas nas áreas de nutrição; dietética; historiografia, antropologia, etnologia e sociologia da alimentação, etc. Heck (2004), que vê na Culinária “o hobby do milênio”, comenta a esse respeito:

O *boom* de livros sobre o assunto “comida” é sem precedentes, e a televisão também descobriu esse filão de sucesso. Livros de receita deixam a prateleira da cozinha e ganham *status* de *coffee-table books*, a literatura toma temas culinários para contar suas histórias, e crônicas prosaicas sobre experiências e memórias de pratos e sabores fazem o maior sucesso. Programas de televisão sobre culinária migram do horário feminino para o horário nobre e ganham formatação de programas populares, e os *chefs* de cozinha se tornam astros da mídia. Heck (2004)

Apesar desse grande leque de possíveis setores de atuação para o tradutor da área, e do aumento crescente da demanda por traduções em todos eles, especialmente junto às editoras, a importância da Culinária no mercado de tradução parece ser ainda tão frágil quanto o próprio status que a área ocupa junto à sociedade brasileira. Note-se, por exemplo, a presença expressiva de edições portuguesas de livros de receitas e obras de referência (como glossários e guias) nas livrarias brasileiras, a má qualidade de algumas obras traduzidas e aqui publicadas, ou a ausência de materiais de referência confiáveis para o tradutor da área (Teixeira, 2004).

Se sairmos da esfera das editoras, deparemos, por exemplo, com cardápios de estabelecimentos respeitáveis traduzidos de tal forma que só poderiam causar riso (ou medo?) no estrangeiro. O jornalista Fritz Utzeri relatou com bom-humor o espanto e a incredulidade que experimentou ao ler o cardápio do hotel em que se hospedou certa vez em Brasília:

Traduzir é difícil (e em geral mal pago). Vejam só o que pode acontecer quando burros informáticos resolvem achar que traduzir é mole e bolam programas de tradução instantâneos que outros, mais burros (ou sovinas) ainda, usam.

Há alguns meses fui a Brasília e hospedei-me num dos hotéis mais finos da cidade. À noite bateu a fome e resolvi consultar o serviço de quarto. Comecei a ler o cardápio, cuidadosamente impresso, e fui ficando assustado. Onde eles teriam arranjado tal tradutor maluco? Tudo era literal. O contrafilé, em português, virou *against filet*. Creme Rachel, um tipo de sopa, foi traduzido para *It cremates Rachel*. “Creme”, substantivo, virou verbo e a ordem era pegar a pobre da Raquel e metê-la num forno até virar cinzas.

Outro prato era à base de "nobre corte de contrafilé". O corte da carne era nobre, mas o "tradutor" achou por bem transformar o adjetivo "nobre" em substantivo e deu num prato de canibal: *nobleman cuts of against filet*, ou seja, "cortes de contrafilé de nobre".

Maminha de alcatra é outro pedaço de carne que pode ser muito traiçoeiro se for traduzido literalmente. Vejam só. Em português o hotel oferece: "delicada peça de maminha grelhada". Maminha foi traduzido, ao pé da letra, como *breast*, que em inglês significa seio, mama ou peito, mas jamais um corte de carne. Imaginem o americano ou inglês horrorizado ao constatar que pode pedir (e comer) *delicate breast pieces grilled*, ou seja "delicados pedaços de seio grelhados".

Mais estranha ainda ficou uma picanha na "manteiga ao café de Paris". Paris, a capital francesa, metamorfoseou-se no verbo "parir", "dar à luz". A coisa ficou assim em inglês (?): *butter coffe of you give birth*, algo que – tentando verter para o português – resultaria em mais ou menos isto: "manteiga café que você pariu".

Já imaginaram o gringo tentando entender frutas da estação (da primavera, verão, etc., *season* em inglês), traduzidas como *fruits of the station*, ou seja, da "estação" (de trem)? Tiras finas de carne resultam em *finer ribbons of meat*. *Ribbon* em inglês é "fita" e não "tira". "Molho de espinafre" passa pelo mesmo processo maluco: o molho (substantivo), *sauce* em inglês, vira *wet*, do verbo "molhar" e resulta em algo que até parece inglês: *I wet of spinach*.

A essa altura o hóspede deve achar que se internou num manicômio e vai ter certeza quando perceber que lhe estão oferecendo *Pitus in coconut*. (Utzeri, 2003).

Se considerarmos com Nery (2006: 49) que o setor de Gastronomia "é o segmento que tem maior participação no PIB do turismo – superando a hotelaria e o transporte", fica difícil entender como a tradução de cardápios não é um dos quesitos observados em seus padrões de excelência em serviço. Vide, por exemplo, alguns pratos que constam do cardápio de um restaurante do Shopping Benfica, em Fortaleza, Ceará, um dos principais pólos turísticos do país:

FILÉ MADEIRA – WOOD STEAK

Filé ao molho Madeira, fritas, arroz à grega e salada verde
Steak with wood sauce, fries, greak [sic] rice and green salad

PEIXE AO MOLHO DE CAMARÃO – FISH SHRIMP SAUCE

Filé de peixe, camarões, arroz branco e purê de batata
Fish steak, shrimps, white rice and smash potato

FILÉ TRINCHADO – SLICED STEAK

Filé bovino, fritas, vinagrete e farofa
Steak Bull, fries, vinegar and farofa

E os brasileiros não são os únicos a difundir idéias errôneas sobre os costumes alimentares do país. No livro *The Ethnic Food Lover's Companion* (Zibart, 2001), o autor, ao falar sobre a culinária brasileira (que reuniu num grupo só com a peruana e a argentina), comete vários disparates, dentre eles, afirmar que: i) o coelho é uma das carnes favoritas, bem como todo tipo de órgãos de animais (p. 408); ii) o abacate e as nozes estão na base da alimentação brasileira (p. 404); iii) a feijoada leva farinha de mandioca e azeite de dendê em seu preparo (p. 410); iv) sopas de frutas são características dos menus brasileiros, assim como sopas que levam leite de coco e nozes/castanhas moídas (p. 409); v) os brasileiros sempre temperam o arroz com tomate, ou então com especiarias e nozes/castanhas picadas (p. 410). Para sedimentar o mal-entendido, à página 415, o autor dá a receita da Feijoada Completa, que mistura o feijão preto, entre outras coisas, com: 1/4 de xícara de xerez seco, vinho do porto ou madeira; 1 1/2 xícaras de suco de laranja; 250g de folhas frescas de couve, mostarda ou espinafre picadas grosseiramente e 2 colheres de sopa de azeite de dendê. O prato é servido com arroz branco, farinha de mandioca torrada, rodela de laranja e abóbora cozida (!).

Mas os problemas de tradução na área do Turismo não se restringem apenas aos escorregões anedóticos da iniciativa privada, cujos profissionais envolvidos são, muitas vezes, mal preparados, apostando sempre na improvisação. Costa (2006), ao estudar as versões para o inglês de material veiculado pela Embratur, conclui:

Com tantas falhas de revisão, traduções que não transmitem a idéia do termo analisado, termos que não são compreendidos pelo leitor-alvo, acreditamos que o material de divulgação fica aquém dos esforços empreendidos nesta área. Isso, a nosso ver, nos associa com uma imagem negativa já bastante difundida nos países desenvolvidos a nosso respeito, de amadorismo e ineficiência. (Costa, 2006: 119).

Em minha experiência pessoal, e com base no que me relatam colegas de profissão e ex-alunos, “amadorismo” parece ser um qualificativo adequado para caracterizar a tradução culinária no Brasil. Aqueles que precisam de um tradutor dessa área não querem gastar muito (afinal, acreditam eles, é uma tradução que não dá trabalho, pois faz parte do dia-a-dia de todos – não é algo preciso como o direito, ou a medicina) e, por esse motivo, repassam as traduções da área para seus tradutores mais inexperientes (reservando os tradutores mais experientes e bem pagos para os temas mais “especializados” e de maior rentabilidade nas vendas).

O tradutor inexperiente, que necessita encontrar seu lugar ao sol no mercado de tradução, aceita, na ingenuidade e meio a contra-gosto, o trabalho mal pago. Traduz, intuitivamente, contando com uma ou outra fonte de referência pouco confiável na Internet, livros sobre o tema (alguns também traduzidos) e seu conhecimento culinário (se os tem). O revisor, pouco acostumado à área, ou apoiando-se em manuais de estilo que foram escritos para textos jornalísticos, e livros não-especializados, faz uma hiper-correção no texto, alterando a terminologia e a sintaxe a seu bel-prazer (já que não existem obras de referência confiáveis em que possa se apoiar nessa tarefa). O livro é publicado. O leitor, ainda em formação e, conseqüentemente, com espírito crítico pouco aguçado, adquire a obra e está criado o problema – esta servirá de guia e fonte de consulta e referência (afinal, foi publicado!) para outras pessoas tão inexperientes quanto as que participaram desse processo. Assim, os erros se propagam.

A primeira vez que traduzi receitas culinárias, que eram parte de um livro maior sobre dieta e menopausa, foi uma espécie de “subcontratação”, por assim dizer – a tradutora responsável pelo trabalho, experiente e ciente das dificuldades impostas por esse tipo de tradução, confiou-me a parte “culinária” do livro. Foi uma ótima oportunidade de exercitar meus conhecimentos teóricos. Infelizmente, quando o livro foi publicado, as decepções foram enormes. Meu nome não constava como tradutora (a editora “esqueceu” de colocar meu nome). Além disso, haviam sido feitas várias alterações no meu texto (algumas delas incorrendo em erros e inadequações), mas não me foi dada a chance de examinar o texto revisado antes da publicação.

Quando o livro foi reeditado, tive que escolher a dedo os problemas “mais graves” para revisar, pois a editora não se dispôs a alterar os fotolitos de todas as páginas com erros por causa de problemas “menores”. Bem, ao menos uma lição foi aprendida: a partir de então, nunca aceitei um trabalho que não me permitisse ler o material antes da publicação (ainda que isso tenha apenas minimizado o problema, já que a última palavra continua sendo, mesmo, a do editor – e isso está, geralmente, nos contratos assinados junto às editoras).

Quando uma outra editora ofereceu o primeiro livro que traduzi sozinha, o valor da lauda proposto era o mesmo pago para uma lauda de tradução “normal” (= texto não-técnico) – R\$14,00 à época. Após trocar alguns e-mails com a equipe responsável, consegui convencê-los de me pagar um valor um pouco maior, reforçando a especificidade do tema. A partir do segundo livro que traduzi para a mesma editora, já

estava recebendo mais que o dobro do que me havia sido oferecido na primeira vez. No entanto, isso parece ser uma exceção.

Uma ex-aluna de tradução com pouca experiência na área relatou-me recentemente que estava traduzindo o quarto livro de uma renomada editora francesa (do inglês para o português, no entanto). Além de falar da dificuldade que estava tendo com a escassez de materiais de consulta, comentou: “De qualquer forma, acho que esse vai ser o último, porque o valor que eles pagam é mesmo simbólico. Não sei quanto as outras editoras pagam, mas R\$10,00 por lauda de 2.000 toques não deve ser! Achei bacana quando me mandaram o primeiro livro com o meu nome, mas a gente tem que valorizar o trabalho, certo?”.

Uma das formas de o tradutor da área valorizar seu trabalho e se fazer reconhecido pelas editoras e agências de tradução é conhecendo a fundo o tema que traduz, de várias perspectivas. Assim, ele será capaz de convencer o contratante de que a Culinária tem suas especificidades como qualquer outra área técnica de tradução, como mostrarei a seguir, além de ser permeada de simbolismos culturais, sociais, econômicos, étnicos, afetivos, para mencionar apenas alguns.

2.2 Breve mapeamento da Culinária

A alimentação é a base da vida, e a Culinária faz parte, de alguma forma, do dia-a-dia de todos os indivíduos e, portanto, de várias esferas da atividade humana. A culinária está presente na cozinha de casa, no bar, no restaurante do hotel, nos estudos acadêmicos de várias áreas, na feira de rua, no supermercado, na Internet, na televisão, no rádio, no jornal, em revistas, na tradição oral e familiar, na literatura, na medicina, etc. Baccin (2003), em sua tese sobre a presença dos italianismos na terminologia gastronômica de São Paulo, aponta várias significações que a alimentação pode adquirir para além da satisfação de uma necessidade básica de nutrição – pode ser um símbolo de afeto e de cuidado, consigo mesmo ou com o outro; um ritual religioso, espiritual; uma forma de recompensa, conquista, agrado, convencimento; um *hobby*; uma forma de recuperar e manter a história individual e/ou familiar; um campo de estudo técnico-científico e acadêmico, etc.

Todos temos algum contato com receitas⁷⁰, o tipo textual culinário por excelência, desde a pessoa que prepara seu Mijojo em casa, seguindo as instruções da

⁷⁰ de que trataremos em detalhe no item 2.3

embalagem, ao engenheiro de alimentos que desenvolveu o pó sabor picanha que vai temperar esse macarrão. Donas-de-casa, empregadas, cozinheiros industriais, chapeiros, garçons, *chefs*, *gourmets*, professores de gastronomia, tradutores, empresas de importação e exportação de alimentos, escritores, telespectadores – todos nos alimentamos e convivemos com os tipos textuais da culinária em algum momento, com maior ou menor envolvimento.

Essa amplitude de alcance do tema e de seus usuários faz com que os tipos textuais orais e escritos pelos quais saberes e sentidos culinários são transmitidos sejam igualmente variados. A seguir, descrevo alguns deles dando ênfase especial à sua presença no mercado de tradução da área.

2.2.1 Embalagens e rótulos de alimentos e bebidas

As embalagens e rótulos de alimentos e bebidas importados ou destinados à exportação devem ser traduzidos. O termo “rótulo” é mais usado para designar uma etiqueta colada ao recipiente que abriga o produto, em geral uma lata ou garrafa; “embalagem” designa um invólucro em que essas informações foram impressas previamente – como caixas de papelão, potes ou sacos plásticos. Uma embalagem / rótulo pode conter vários subtipos textuais:

2.2.1.1 Informações nutricionais

A parte do rótulo / embalagem que trata das informações nutricionais apresenta, em forma de tabela (vide Fig. 4), uma lista de nutrientes (geralmente: carboidratos, proteínas, gorduras – totais, saturadas e trans –, colesterol, fibra alimentar, sódio, açúcar e vitaminas) e, com relação a uma determinada medida (como 15g / 1 1/2 colheres de sopa), seu peso relativo em gramas⁷¹, na primeira coluna, e numa segunda coluna, a porcentagem relativa a uma dieta de X calorias diárias (atualmente, 2.000 kcal / 8.400 kJ).

Observe-se que a tradução desse texto deve seguir as normas padronizadas do país de destino para esse tipo de texto (consulte, p. ex. a ANVISA⁷² no Brasil, e o FDA⁷³ nos Estados Unidos). Exemplos:

⁷¹ o peso nessa tabela é sempre dado em gramas e/ou em porcentagens relativas à medida padrão (15 g / 1 1/2 xícaras, neste exemplo), mesmo nos países que usam um sistema métrico diferente.

⁷² <http://www.anvisa.gov.br/rotulo/>

⁷³ <http://www.intertek-cb.com/food/Foodtestlabel.shtml>

INFORMAÇÃO NUTRICIONAL		
Porção de 15 g (1 e 1/2 colher de sopa)		
Quantidade por Porção		% VD (*)
Valor energético	38 kcal = 160 kJ	2 %
Carboidratos	0 g	0 %
Proteínas	0 g	0 %
Gorduras totais	3,8 g	7 %
Gorduras saturadas	2,4 g	11 %
Gorduras trans	não contém	**
Fibra alimentar	0 g	0 %
Sódio	7,8 mg	0 %

* %Valores Diários de referência com base em uma dieta de 2.000 kcal ou 8.400 kJ. Seus valores diários podem ser maiores ou menores dependendo de suas necessidades energéticas. ** VD não estabelecido.

NESTLÉ® LA LECHERA® Sweetened Condensed Milk 14.1 OZ.	
Serving Size 2 Tbsp (30 mL) Servings Per Container about 10	
Amount Per Serving	
Calories	130
Calories from Fat	30
% Daily Value*	
Total Fat 3g	5%
Saturated Fat 2g	10%
Trans Fat 0g	
Cholesterol 10mg	3%
Sodium 45mg	2%
Total Carbohydrates 22g	7%
Sugars 22g	
Protein 3g	
Vitamin A	2%
Calcium	10%

INGREDIENTS: MILK, SUGAR

Not a significant source of dietary fiber, vitamin C, and iron.
*Percent Daily Values are based on a 2,000 calorie diet.

Figura 4: informações nutricionais do creme de leite Nestlé® em português e inglês

2.2.1.2 Lista de ingredientes

Nessa lista são enumerados os ingredientes que entram na composição do produto: desde a matéria-prima natural (frutas, carnes, óleos, etc.) aos conservantes, aditivos, corantes, flavorizantes, aromatizantes, melhoradores, etc. Em geral é dado destaque para a presença de ingredientes ou substâncias que possam causar alergias, como glúten, castanhas e nozes, leite, ovos crus, etc. Assim como no caso das informações nutricionais, esse texto deve respeitar normas federais e internacionais pré-determinadas na descrição da composição do produto. Exemplo:

Figura 5: Lista de ingredientes – pacote de granola

2.2.1.3 Instruções de preparo, sugestão de consumo e receitas

Quando o produto não está pronto para o consumo imediato, necessitando ser cozido, esquentado, resfriado, misturado a outros ingredientes, etc., há um texto na embalagem explicando como isso deve ser feito (vide quadro à esquerda na Fig. 6). Em alguns produtos, além dessa explicação básica (p. ex.: como esquentar a sopa), são oferecidas sugestões de consumo (p. ex.: sirva acompanhado de torradas/polvilhado com queijo) e, em alguns casos, são oferecidas receitas inteiras (vide quadro à direita na Fig. 6).

Figura 6: instruções de preparo e receitas – sopa Campbells®

2.2.1.4 Nome do produto, descritivo, foto e outros elementos

Há outros elementos da embalagem / rótulo que o tradutor pode ser solicitado a traduzir, tais como: i) um descritivo do produto (por exemplo: "Italy's #1 brand of pasta", na embalagem das massas Barilla; "60 anos de qualidade", na embalagem da farinha de mandioca Deusa); ii) instruções de como abrir a embalagem (p. ex: "press and lift up", "abra aqui"); iii) informações sobre o fabricante / importador e os dados para contato com o setor de atendimento ao cliente; iv) o peso do produto (líquido / drenado – no caso de alimentos conservados em calda, salmoura, etc.).

O nome do produto, as cores da embalagem e outros aspectos visuais em geral são trabalhados pelo setor de marketing. Mas se o tradutor achar conveniente intervir, seja por questões culturais, seja por questões lingüísticas, pode fazer sugestões no momento da tradução.

2.2.1.5 Bebidas alcoólicas: um caso à parte

A tradução de rótulos de bebidas alcoólicas importadas ou destinadas à exportação, assim como no caso dos alimentos, deve seguir um padrão terminológico rígido, ainda que não seja imutável.

A terminologia relacionada aos vinhos, por exemplo, que pode aparecer tanto em rótulos de produtos quanto em livros e materiais especializados sobre o assunto, apesar de padronizada, é subjetiva e pouco transparente, como salienta Dória:

Difícilmente um neófito é capaz de compreender o que se pretende comunicar quando se diz de um vinho que ele é “marcante no ataque, frutado, com notas de baunilha, tabaco e couro, encorpado, persistente, complexo e com boa lágrima”. Mesmo os iniciados, que se dizem capazes de compreender o código, não o são para explicá-lo convincentemente a terceiros. (Dória, 2006: 173).

Segundo esse autor, em 1976 foi feita uma apuração que identificou 611 termos (i. e., descritivos) do vocabulário enológico. Essa terminologia, na opinião de especialistas, é adequada para descrever a percepção sensorial e fisiológica e o tratamento dos estímulos recebidos, mas falha ao tentar traduzir / comunicar a informação sensorial por meio de analogias – um problema muitas vezes relacionado a diferenças do mundo físico e da cultura de regiões geográficas distintas:

Tome-se o exemplo da prova olfativa, a mais complexa e controvertida na análise dos vinhos. As impressões olfativas dos degustadores são, em geral, traduzidas em analogias extraídas de todas as fontes que encontram em seu entorno. Assim, quando o vinho se mundializa não é mais possível pretender que alguém que degusta um vinho em São Paulo saiba o que quer dizer o enólogo quando se refere a “aroma de raposa molhada”. (Dória, 2006: 175)

O autor cita o caso de vinhos argentinos que, em recente ofensiva comercial na Inglaterra, qualificaram sua produção como “vinhos de ritmo” – uma alusão ao tango, uma das representações culturais do país no exterior.

A tentativa de vincular as características de uma bebida alcoólica a uma identidade nacional ou regional no mercado de exportação tem até mesmo implicações jurídicas, no caso de bebidas típicas (feitas a partir de ingredientes locais, ou seguindo

métodos diferenciados de fabricação). Assim como as vinícolas possuem hoje regras para classificar os vinhos de acordo com Denominações de Origem Controlada (DOC), bebidas nacionais, como a cachaça, têm trilhado caminhos no mercado internacional em busca do reconhecimento de bebida típica brasileira.

Costa (2006) analisa a tradução da palavra “cachaça” em dicionários bilíngües e num *corpus* paralelo⁷⁴ composto de textos de divulgação cultural sobre o Brasil escritos em português, e sua versão correspondente em inglês, veiculados por órgãos de turismo, revistas de bordo e *websites*. A pesquisadora observa que, na maioria dos casos, “cachaça” é traduzido como “cachaça”, mas há muitos casos em que o aposto colocado entre parênteses logo depois da palavra identifica a bebida como um subtipo de outras bebidas mais difundidas na língua inglesa – “typical Brazilian sugar cane **rum**”, “sugarcane **brandy**”, “raw white **rum**”, “sugar-cane **rum**”, “sugar cane **spirits**”.

Costa conclui que a variação é prejudicial não apenas para a criação de uma identidade do produto junto ao público externo – “Como é que o estrangeiro vai entender que se trata de uma coisa típica do Brasil se a cada leitura ele a encontra de uma forma diferente?” (*op. cit.*: 207) –, mas também do ponto de vista jurídico e comercial:

A pedido dos produtores de cachaça, o Ministério das Relações Exteriores vem negociando arduamente junto aos órgãos internacionais aduaneiros para registrar a cachaça como bebida alcoólica diferenciada das outras bebidas produzidas a partir da cana-de-açúcar. Os esforços do Itamaraty têm sido no sentido de descaracterizar a cachaça como rum (produto sobre o qual incidem altas taxas tarifárias, principalmente nos EUA, onde a cachaça só entra se apresentar um rótulo de “*Brazilian rum*”) e criar uma categoria aduaneira própria para nossa bebida de forma a tornar o nome cachaça conhecido. Há, ainda, uma solicitação de conferir proteção de indicação geográfica à cachaça, garantindo-a como produto exclusivo do Brasil, como já é feito com outras bebidas como o vinho do Porto, por exemplo, já que há cachaças sendo fabricadas em outros países como a Colômbia. (Costa, 2006: 207)

O que se pode perceber, portanto, é que o tradutor que optar por trabalhar nessa área deverá aprofundar-se no assunto, por meio de cursos, leituras, participação de eventos da área, etc. e, se possível, contar com uma consultoria especializada nos casos que envolvem implicações jurídicas como as que foram apresentadas aqui.

⁷⁴ Corpus MIME, com 470 pares de textos que totalizam cerca de 720 mil palavras – 356.000 em português e 362.000 em inglês (Costa, 2006: 201)

2.2.2 Cardápios de restaurantes

Como já exemplificado anteriormente, os problemas de tradução em cardápios de restaurantes e hotéis podem ser anedóticos e, algumas vezes, hilariantes. Mas o outro lado da situação preocupa: um cardápio mal traduzido passa para o turista estrangeiro uma imagem negativa do estabelecimento e, quando o problema é generalizado, da região e do país visitados. A consciência desse problema levou a Abrasel – Associação Brasileira de Bares e Restaurantes – a oferecer, a partir de 2008, um serviço de tradução de cardápios para seus associados, que pode ser solicitado, pelos associados, junto ao *site* da entidade⁷⁵:

No ano passado, o país recebeu 5 milhões de estrangeiros, segundo dados da Empresa Brasileira de Turismo (Embratur). Para recebê-los melhor este ano, a Associação Brasileira de Bares e Restaurantes (Abrasel) oferece a tradução de cardápios. Um menu em outras línguas facilita o trabalho do garçom que poderá fazer uma introdução simples no idioma estrangeiro sem cometer gafes, destaca o presidente executivo da Abrasel, Paulo Solmucci Júnior. (Abrasel, 2008)

A associação espera atender 800 estabelecimentos de todo o país numa fase inicial, e cerca de 2 mil numa segunda etapa. Os 1.534 restaurantes que participaram do Movimento Sabor Brasil – festival gastronômico do país que ocorre em 170 cidades de diversas regiões turísticas – terão prioridade na tradução. Além disso, a Abrasel está desenvolvendo um projeto em parceria com Ministério do Turismo que prevê o treinamento à distância de garçons e gerentes em língua inglesa e espanhola por meio de vídeo-aulas.

Cardápios⁷⁶ necessitam de traduções tanto para os nomes de pratos e bebidas (sejam eles tradicionais ou exclusivos do estabelecimento em questão), quanto para seus descritores. A tradução pode ser intercalada com o cardápio original, ou podem ser criadas versões separadas, uma para cada língua, como no exemplo da Fig. 7, que mantém os nomes originais dos pratos e traduz a explicação de sua composição na língua estrangeira.

⁷⁵ http://www.abrasel.com.br/index.php/servicos/faca_a_traducao_do_seu_cardapio_gratuitamente_e_aumentando_suas_vendas/

⁷⁶ “*Menu* vem do latim *minutus*, ‘resumido’, ‘diminuto’. Não deve ser confundido com o cardápio ou ementa (*carte*), que é a lista de tudo que um restaurante oferece. Um cardápio pode conter vários menus ou sugestões de refeições com seus diferentes serviços e etapas.” (Franco, 2001: 204)

Figura 7: exemplo de menu em inglês – restaurante *Tordesilhas*, SP

2.2.3 Interpretação para turismo gastronômico

Um novo nicho no mercado de tradução culinária consiste na oferta de serviço de interpretação junto a órgãos e empresas voltadas ao turismo gastronômico, ou mesmo em restaurantes e hotéis turísticos com grande volume de clientes estrangeiros. O profissional acompanha o turista estrangeiro em suas visitas a restaurantes, feiras, mercados e eventos relacionados à Culinária regional e nacional explicando cardápios, descrevendo produtos e preparações típicas nacionais, etc.

(...) há quem prefira a praticidade de desfrutar um roteiro turístico com a presença de um tradutor. É o caso de um grupo de franceses que conheceu Brasília, no último dia 2, acompanhados do guia Paulo Srdic, profissional da empresa de turismo receptivo Prestheza. (Abrasel, 2008)

É possível trabalhar também junto a agências que organizam eventos internacionais na área de Gastronomia para fazer interpretação de aulas, palestras, workshops, tradução de material a ser distribuído aos participantes, etc. O tradutor que optar por atuar nessa área deve receber treinamento especial como intérprete e ter um conhecimento aprofundado e atualizado da culinária brasileira e internacional em seus aspectos culturais, históricos, etnográficos, sociais, etc.

2.2.4 Legendagem de material audiovisual sobre culinária

Os programas sobre culinária e turismo gastronômico, com as demandas do sistema de TV a cabo implantado no Brasil no início dos anos 1990, passaram a ser um mercado de atuação possível para o tradutor da área. De lá para cá, vários programas de culinária em língua inglesa passaram a ser exibidos na TV por assinatura, como o de Martha Stewart, Jamie Oliver, Claude Troisgros, Nigela, etc.

Além de programas sobre culinária, o tradutor pode atuar na tradução e versão de filmes sobre o tema, bastante recorrentes na filmografia dos últimos anos (p. ex.: *Estômago*, *Ratatouille*, *Super size me*, *Mostly Martha*), e de DVDs educativos, videoaulas, programas sobre turismo gastronômico e culinária nacional / regional, etc.

2.2.5 Tradução de material multimídia

Há centenas de *websites* na Internet dedicados exclusivamente à Culinária e à Gastronomia, e outros tantos em que os temas figuram entre os links da página, oferecendo receitas, informações sobre alimentação saudável, etc. O profissional que optar por trabalhar nessa área pode traduzir e verter, por exemplo, o conteúdo de *sites* institucionais de empresas do setor de alimentos e bebidas, de restaurantes, blogs sobre culinária, *sites* de *chefs*, escolas de culinária, empresas de *catering*, pousadas, revistas e jornais especializados em Culinária e/ou Nutrição, *sites* de guias turísticos, de associações de restaurantes e bares, etc.

2.2.6 Tradução de livros e outros materiais impressos

Além dos tradicionais livros de receita, o tradutor pode trabalhar traduzindo e/ou vertendo outros tipos de livros e materiais impressos que incluem a temática culinária, como revistas especializadas ou que tenham uma seção de Culinária, informativos e jornais de associações de classe (bares e restaurantes, padarias, pousadas, etc.), material promocional de alimentos e bebidas (folhetos com receitas de produtos recém-lançados, em promoção, produtos sazonais, etc.), material didático de cursos de treinamento em nível técnico e superior, revistas de bordo de empresas aéreas, entre outros.

2.3 Receita: o texto culinário por excelência

A receita é o tipo textual mais difundido e representativo da Culinária, tendo sofrido enormes mudanças em seu formato ao longo dos séculos, desde que deixou de ser somente uma tradição oral.

Os primeiros compêndios culinários europeus de que se tem registro datam do século XIII, mas há de se considerar que esses manuscritos, por serem usados na cozinha, estavam mais sujeitos à degradação, além de seu status como documento histórico ser relativamente recente (Couto, 2007: 21-22).

A impressão de livros de culinária – um dos temas pioneiros no mercado editorial da Europa (Franco, 2001: 142) – teve início no século XVI, mas o primeiro livro em língua portuguesa, *Arte de Cozinha*, de Domingos Rodrigues, foi publicado em Portugal bem depois, em 1.680. A obra marca o momento de transição que houve durante a Renascença entre a cozinha medieval da Baixa Idade Média (séculos XIV a XV), caracterizada pela quantidade e pelo uso excessivo de especiarias, e a nova cozinha "à francesa", que prezava mais a qualidade, e cujas características se assemelham mais às da cozinha como é conhecida hoje.

A publicação do português *Arte de cozinha*, portanto, surge num contexto de transformação profunda da cozinha cortesã européia: práticas alimentares comuns no medievo, como longos cozimentos, uso de especiarias em profusão e de alguns (poucos) vegetais muito bem cozidos foram cedendo lugar à nova proposta. Esta, em termos gerais, pregava a utilização de especiarias em menos quantidade, incluía no cardápio das elites frutas e vegetais crus e transportava o açúcar – utilizado na maioria dos pratos até então – para o final da refeição. (Couto, 2007: 39)

Essa nova cozinha passou a ser difundida em publicações como *Grand cuisinier de toute cuisine*, de autoria coletiva e publicado pela primeira vez em 1540, e *Le cuisinier françois*, de François Pierre de La Verenne, de 1651.

O livro de cozinha da infanta d. Maria de Portugal, manuscrito levado por D. Maria de Portugal para Parma, em 1565, por ocasião de seu casamento com o duque de Parma, Placêncio e Castro, pode ser considerado um típico representante dessa cozinha medieval. Reúne 61 receitas – 26 de “manjares de carne”, 4 de “manjares de ovos”, 7 de “manjares de leite” e 24 de “coisas de conservas”. As preparações têm por base a combinação de três grupos de alimentos, que representam os três sabores

fundamentais que caracterizam as preparações medievais: as especiarias (sabor forte), o açúcar (sabor doce) e o vinagre (sabor ácido) (Couto, 2007: 41-42).

O primeiro livro de cozinha publicado no Brasil de que se tem notícia intitula-se *Cozinheiro Imperial*, e é de 1840. Assinado pelo *chef* R.C.M., têm por principal objetivo “oferecer ao Brasil um manual da ‘ciência culinária’ que equipare o país às nações européias” (Couto, 2007: 120). Sua sétima edição, de 1877, reuniu 1.200 receitas, além de técnicas de cozinha (cozinhar em banho-maria, trincar carnes diversas), dicas de organização de banquetes (arrumação da mesa, como servir), procedimentos de higiene e manutenção de utensílios de cozinha, etc.

Assim como os livros portugueses em que se baseia (*Arte de Cozinha* e *Cozinheiro moderno*, escrito por Lucas Rigaud em 1780), divide as receitas de acordo com os alimentos que entram em sua preparação. Do ponto de vista textual, a preferência pelo verbo no modo imperativo já se manifesta, mas as receitas ainda não separavam os ingredientes do modo de preparo (Abbade, 2005), mimetizando a tradição oral – apenas algumas receitas mantêm esse formato atualmente, como se pode ver, a seguir, no item 2.3.2.

De lá para cá, muitos livros de culinária foram publicados e traduzidos em todo o mundo, e a receita sofreu diversas alterações em sua forma de organização, conteúdo e apresentação visual, acompanhando as mudanças da sociedade como um todo e a evolução da culinária ocorrida com o intercâmbio de ingredientes e técnicas proporcionado pela globalização. A seguir, proponho alguns critérios para descrever as principais características desse tipo textual atualmente no inglês e no português brasileiro do ponto de vista da tradução técnica.

2.3.1 Receitas quanto ao meio de divulgação

Um primeiro critério que poderia ser usado para distinguir as receitas quanto ao seu meio de divulgação seria separá-las entre orais e escritas.

As receitas orais podem ocorrer em diversas esferas de interação social, privadas ou públicas, entre elas: conversações entre amigos / familiares / colegas de trabalho, entre garçom / *maître-d'hôtel* e cliente de restaurante ou membros da equipe de cozinha, entre profissionais que oferecem degustação de produtos em supermercados, etc. e clientes; programas de rádio; programas televisivos;

atendimento a cliente por telefone (oferecido por empresas do ramo de alimentos e bebidas); aulas e workshops de Culinária.

As receitas escritas podem ser subdivididas entre publicadas e não-publicadas. As não-publicadas figuram em documentos como cadernos de receitas e materiais didáticos usados em cursos de treinamento e escolas de Culinária, conversações, aulas. As publicadas podem ser divulgadas em meio audiovisual (quando uma receita escrita é exibida num programa de televisão, por exemplo), mas principalmente em meio impresso e eletrônico.

As receitas publicadas em meio-eletrônico podem ocorrer em *sites* dedicados exclusivamente à culinária (p. ex.: Basílico⁷⁷, Batuque na Cozinha⁷⁸), em *sites* diversos que tenham uma seção dedicada à culinária (p. ex.: Terra culinária⁷⁹), blogs (p. ex.: Dadivosa⁸⁰, Mixirica⁸¹ – ambos fizeram tanto sucesso que agora são *sites*, mas com o formato de blog), *sites* institucionais diversos (p. ex.: Nestlé⁸², Supermercados Pão de Açúcar⁸³, link Gastronomia da seção Vida Provinciana do jornal eletrônico *A Província*, de Piracicaba⁸⁴).

As receitas publicadas em papel estão em: livros; jornais, seja em cadernos especiais (p. ex.: Paladar, do *Jornal O Estado de S. Paulo*), seja em seções de suplementos voltados ao Turismo, Lazer, etc. (coluna Culinária do *Jornal de Uberaba*⁸⁵); em revistas dedicadas ao segmento (p. ex.: *Gula*, *Cláudia Cozinha*, *Padaria Moderna*) e em outras revistas diversas (p. ex.: seção Paladar da *Revista Contigo*, seção Prazer à mesa da *Revista Uma*); em embalagens e rótulos de produtos alimentícios; em material de divulgação de produtos alimentícios; em jornais / panfletos de supermercados; em calendários (p. ex.: a folhinha do Sagrado Coração de Jesus), entre outros.

Na presente pesquisa trabalho apenas com receitas escritas e publicadas em meio impresso e/ou eletrônico, cujas características passo a detalhar.

⁷⁷ <http://basilico.uol.com.br/>

⁷⁸ <http://batuquenacozinha.oi.com.br/>

⁷⁹ <http://culinaria.terra.com.br/>

⁸⁰ <http://www.dadivosa.org/>

⁸¹ <http://mixirica.uol.com.br/>

⁸² <http://www.nestle.com.br/Cozinha/Home/Default.aspx>

⁸³ <http://www.paodeacucar.com.br/culinaria.asp?catCn=85606177-996C-4B81-B33F-F5144273094F>

⁸⁴ <http://www.aprovincia.com/padrao.aspx?texto.aspx?idContent=45365&idContentSection=693>

⁸⁵ <http://www.jornaldeuberaba.com.br/?MENU=Colunas&SUBMENU=culinaria>

2.3.2 Receitas quanto à organização e à aparência

Como já comentei anteriormente, há duas formas básicas de apresentação de uma receita escrita. A mais difundida atualmente divide o conteúdo em seções, como ingredientes e modo de fazer (mais detalhes no item 2.3.3). Há também a forma “corrida”, que mimetiza a tradição oral ao mencionar os ingredientes, sua quantidade, etc. conforme são utilizados no procedimento. Este tipo de receita parece estar se tornando cada vez mais raro na forma escrita, ao menos no português brasileiro⁸⁶:

:: MEXIDINHO DE CARNE E BRÓCOLIS

Coloque 2 colheres (chá) de óleo na panela e aqueça. Adicione 250 gramas de carne bovina picada em tirinhas e mexa por 2 minutos. Junte 2 xícaras de brócolis e couve-flor, 1 cebola cortada em rodela finas, 1 pimentão cortado em quadradinhos, 1/2 xícara de milho, 1 colher (sopa) de alho picado e misture por 2 minutos. Acrescente 1/2 xícara de castanhas de caju e 1 colher (sopa) de molho shoyu light (com menos sal). Mexa por mais 1 minuto.⁸⁷

Uma receita típica de um livro de receitas ou *website* especializado em Culinária em geral apresenta os ingredientes em forma de lista, seguidos das etapas do preparo – que podem formar um bloco só de texto corrido, ou subdividirem-se em parágrafos ou itens, numerados ou não. É comum a apresentação de uma foto do prato final e, nos materiais destinados ao usuário inexperiente, fotos das etapas do preparo (vide Fig. 8). Alguns livros voltados para *chefs* e usuários com nível mais avançado de especialização têm dispensado ou reduzido bastante o uso de fotografias e/ou ilustrações (p. ex.: *How to eat: The pleasures and principles of good food*, de Nigella Lawson, 2000).

⁸⁶ encontramos várias ocorrências de receitas desse tipo durante a compilação de um *corpus* multivarietal contrastivo (Teixeira, 2004; Tagnin e Teixeira 2004a e 2004b)

⁸⁷ http://mdemulher.abril.com.br/anamaria/edicoes/565/culinaria/anamariamateria_245901.shtml [acesso em 22/8/2008]

Espaguete à carbonara

No carbonara, é o calor do espaguete recém-escorrido que cozinha ligeiramente o molho cremoso de gemas. Este prato fica pronto num instante, então tenha todos os ingredientes à mão antes de começar.

TEMPO DE PREPARO: 10 minutos
TEMPO DE COZIMENTO: 5 minutos
RENDIMENTO: 4 porções

4 gemas
 2 ovos
 150 ml de creme de leite fresco
 50 g de queijo pamesão ralado
 2 colheres (sopa) de azeite
 100 g de panceta ou bacon em fatias cortados em tiras finas
 2 dentes de alho amassados
 400 g de espaguete fresco
 sal e pimenta-do-reino

- 1 Bata junto as gemas, os ovos inteiros, o creme de leite, o queijo e bastante pimenta-do-reino.
- 2 Aqueça o azeite em uma frigideira grande e frite a panceta (ou bacon) por 3-4 minutos, ou até que fique dourada e crocante. Junte o alho e deixe cozinhar por mais 1 minuto.
- 3 Enquanto isso, ferva água com sal em uma panela grande, junte o espaguete e cozinhe por 2 minutos, ou até que esteja macio.
- 4 Escorra e transfira imediatamente para a frigideira. Desligue o fogo e junte o molho de gemas, mexendo, até que os ovos estejam ligeiramente cozidos. Sirva imediatamente. Se o calor da massa não for suficiente para cozinhar o molho de gemas, acenda o fogo e aqueça rapidamente, mexendo sem parar.

46 *Simplesmente molhos*

Espaguete à carbonara

Figura 8: exemplo de receita com fotos do passo a passo e do prato final⁸⁸

2.3.3 Componentes estruturais e lingüísticos da receita

Para que um texto seja reconhecido como exemplar de um determinado gênero⁸⁹ numa dada cultura é preciso que ele possua características macro e microtextuais que permitam ao leitor identificá-lo como tal. Conforme demonstrado em trabalhos anteriores (Teixeira 2004; Tagnin e Teixeira 2004a e 2004b), as

⁸⁸ Farrow, Joanna (2007). *Massas: 80 das melhores receitas*. Tradução de Elisa D. Teixeira. São Paulo: Manole.

⁸⁹ A exemplo de Colina, faremos aqui uma distinção entre “tipo textual” e “gênero textual”: “Genre ascription is at least in part determined by the social context in which a text exists, resulting in generic categories such as novel, poem, editorial, joke, ad, recipe, news broadcast. (...) Text types refer to the main rhetorical purpose of a text, i.e. argumentative, informative, expressive, persuasive, descriptive (...)” [O gênero é, ao menos em parte, determinado pelo contexto social em que o texto se insere, resultando em categorias como romance, poema, editorial, piada, anúncio, receita, noticiário. (...) Os tipos textuais referem-se ao propósito retórico principal de um texto, ou seja, argumentativo, informativo, expressivo, persuasivo, descritivo (...)] (COLINA, 1997: 336)

características textuais peculiares ao gênero “receita culinária” mudam não só de língua para língua, mas também de cultura para cultura. As receitas brasileiras, quanto à extensão, nível de detalhamento e preferência por medidas caseiras, por exemplo, têm mais em comum com as receitas americanas do que com as portuguesas, cujo estilo assemelha-se mais ao das receitas britânicas (*id. ibid.*). Kittredge (1982: 113) comenta, a esse respeito: “Since written recipes are generated spontaneously in most languages with written tradition, cultural borrowing may be less of an influence on sublanguage style than in highly technical sublanguages”⁹⁰.

Quanto ao tipo textual (isto é, seu propósito retórico), as receitas são predominantemente “instrucionais” (Colina, 1997), ou “operativas”, na terminologia de Nord (1997). Orais ou escritas, descrevem um procedimento, muitas vezes decomponível em várias etapas sucessivas e complementares, que deve ser seguido com vistas a se obter um determinado resultado, seja na receita culinária, seja na médica, de felicidade, da eterna juventude, de beleza, de sucesso, etc. Kittredge (1982: 113) chega mesmo a classificar as receitas como um tipo de manual técnico, referindo-se ao modo de preparo como “assembly instructions” (= instruções de montagem). Ele justifica essa classificação com base na semelhança da linguagem: “The procedural sentences of the recipes show a pattern of deletion for definite object noun phrases (NPs) which is typical of many technical manuals.”⁹¹ (*op. cit.*: 114).

Para Nord (1997), nos textos do tipo “operativo”, tanto o conteúdo quanto a forma estão subordinados ao efeito extralingüístico pretendido. Isso traz implicações diretas para a postura a ser adotada pelo tradutor desses textos:

The translating of operative texts into operative texts should be guided by the overall aim of bringing about the same reaction in the audience, although this might involve changing the content and/or stylistic features of the original.⁹² (Nord, 1997: 38)

Tanto a tipologia quanto o gênero textual da receita devem ser observados pelo tradutor da área, levando-se em conta as condições de recepção do texto traduzido –

⁹⁰ Uma vez que as receitas escritas são geradas espontaneamente na maioria das línguas com tradição escrita, empréstimos culturais podem ser menos influentes no estilo dessa linguagem especializada do que nas que tem um grau mais alto de especificidade técnica.

⁹¹ As sentenças do procedimento de uma receita exibem um padrão de apagamento dos sintagmas objetivos definidos que é típico de muitos manuais técnicos.

⁹² A tradução de textos operativos para textos operativos deve ser conduzida pelo objetivo principal de provocar o mesmo tipo de reação no público, embora isso possa envolver mudanças de conteúdo e/ou das características estilísticas do original.

público-alvo, meio de divulgação, características macrotextuais, recursos icônicos, etc. O caráter “operativo/instrucional” deve conduzir as escolhas durante o fazer tradutório pois, ao contrário do que geralmente se preza na teoria e na prática da tradução de outros tipos textuais, a “fidelidade” é preterida, nesses textos, em favor da “funcionalidade”. Salvo raras exceções (compêndios históricos, receitas em textos literários, anedóticos e humorísticos, por exemplo), a receita traduzida deve, acima de tudo, ser exeqüível – ainda que isso represente, como salienta Nord no trecho citado, operar mudanças no conteúdo e/ou na forma do texto original (cabe ao tradutor identificar os casos esporádicos em que essa característica é menos importante).

Quanto à macroestrutura, os itens essenciais de uma receita são: título (o resultado que se pretende obter), ingredientes (os elementos que é preciso transformar para obter esses resultados) e procedimento (como se deve operar esses elementos). Mas as receitas culinárias podem apresentar também outros itens, seja para facilitar seu entendimento e execução por parte do usuário, seja para enriquecer seu conteúdo informativo, cultural, nutricional, histórico, etc. Do ponto de vista microtextual, as receitas se caracterizam pelo uso de estruturas textuais, vocabulário e sintaxe típicos da área.

No que se segue, proponho uma caracterização das receitas a partir do trabalho de alguns autores, em especial Ostmann e Baker (2001) e Jacob (2005), e de pesquisas feitas na Internet e no *corpus* compilado para esta pesquisa (vide item 3.5).

2.3.3.1 Elementos obrigatórios

A. Título

O título é de suma importância para uma receita – é o primeiro contato que o leitor tem com o texto, informando-lhe a natureza da preparação e/ou instigando-o a preparar a receita. Uma busca na Internet com o auxílio da ferramenta Google revelou algumas características comuns a títulos de receitas brasileiras.

Os títulos podem, por exemplo, descrever os ingredientes usados na receita (“Pêras ao chocolate”, “Truta com laranja”, “Filé de frango com cogumelos”); a técnica empregada no preparo (“Filé grelhado com ervas finas e manteiga”, “Lombo assado”, “Peixe no vapor”, “Batata-doce frita”); indicar a variação de ingredientes numa preparação já conhecida (“Torta de cebola”, “Suflê de milho verde”, “Bolo de cenoura”, “Omelete de espinafre com queijo”) ou a variação de uma receita típica (“Carpaccio de

abobrinha”, “Moqueca de caju”, “Goulash de champignon”, “Tabule de frutas”, “Bobó de berinjela”). Pode, ainda, usar expressões apreciativas (“Torta deliciosa de bacalhau”, “Bolo saboroso de café”, “Bolo delícia de pêssego”) e, em alguns casos, ser de sentido não transparente (“Bife a cavalo”, “Baião de dois”, “Vaca atolada”, “Escondidinho”, “Arrumadinho”, “Nega Maluca Branca” [sic], “Meia de seda”, “Baba de moça”, “Floresta Negra”, “Sonho”, “Toalha felpuda”), ou qualquer combinação dessas características (ex: “Salada quero mais”, “Bolo de chocolate peteleco”, “Torta folhada fantasia”, “Torta de frango econômica”, “Pizza de liquidificador”, “Surpresa de ovos”).

Ostmann e Baker (2001) e Jacob (2005), ao descreverem as qualidades de um bom título de receita em língua inglesa, reforçam que este deve ser, acima de tudo, descritivo e convidativo:

Recipe titles should be simple, descriptive, informative, and inviting. You don't want them to read like a menu item in a restaurant. And you don't want them to be vague, so readers can't envision the dish.⁹³ (Jacob, 2005: 171)

The recipe title or name should be descriptive as well as interesting. It should attract the attention of the reader while providing information about the recipe. Avoid the use of brand names and flowery or cute words.⁹⁴ (Ostmann e Baker, 2001: 15)

O título em geral não comporta o uso de notas explicativas para ingredientes ou termos pouco conhecidos (como, por exemplo, “Omelete de pimentão e chalotas (cebolas pequenas)”⁹⁵) – isso é feito na lista de ingredientes, nos comentários introdutórios ou em notas de fim / rodapé.

B. Ingredientes

Os ingredientes, na maioria das receitas, aparecem separados, logo após o título (e os comentários, quando há algum), em forma de lista disposta em uma ou duas

⁹³ O título da receita deve ser simples, descritivo, informativo e convidativo. Não é para ficar parecido com um item de menu de restaurante. E também não é para ser vago, de forma que o leitor não possa imaginar o prato.

⁹⁴ O título ou nome da receita deve ser descritivo e também interessante. Deve atrair a atenção do leitor e, ao mesmo tempo, fornecer informações sobre a receita. Evite o uso de marcas comerciais e palavras muito rebuscadas ou modismos.

⁹⁵ Título de receita retirado de: AGATSTON, Arthur. *A dieta de South Beach. Um plano delicioso e garantido para perder peso de forma rápida e saudável*. Trad. Ana Beatriz Rodrigues. Rio de Janeiro, Sextante: 2003.

colunas. Quando o prato final requer o preparo de receitas intermediárias (p. ex.: massa, recheio e cobertura numa receita de bolo), os ingredientes são listados separadamente para cada uma delas. Cada ingrediente ocupa uma linha (que não deve ter pontuação ao final) e, em português, é precedido de uma unidade de medida que determina sua quantidade (xícaras, colheres; kg, g, l, ml; maço, ramo, talo, dente, lata, pitada, etc.) seguida da preposição “de” (p. ex.: “1 kg *de* carne moída”, “3 colheres (sopa) *de* mel”). Em inglês, a preposição entre a unidade de medida e o ingrediente é sempre omitida (p. ex.: “2 cups flour”, “3 fl oz chicken stock” “4 large carrots”).

As receitas brasileiras e americanas destinadas ao público geral dão preferência às medidas caseiras (Tagnin e Teixeira, 2004b). Em português, as receitas podem apresentar a medida caseira ou da embalagem e seu correspondente em gramas ou mililitros lado a lado, separados por uma barra ou entre parênteses (p. ex.: “250 ml / 1 xíc. (chá) de leite”, “1 envelope (12 g) de gelatina em pó sem sabor vermelha”). Em inglês, as medidas equivalentes no sistema métrico internacional (bem como a temperatura em graus Celsius, no modo de preparo) podem aparecer entre parênteses (p. ex.: “1 lb (500 g) bow tie pasta”, “7/8 cup (7 oz / 200 g) lard”). Atualmente, somente os Estados Unidos publicam alguns poucos livros em que as medidas são dadas somente no sistema métrico norte-americano / imperial. Ao fazer uma versão para o inglês, convém o tradutor oferecer sempre o equivalente no sistema métrico internacional, para que o livro possa ser distribuído em todos os mercados editoriais.

Em língua portuguesa, logo após o ingrediente pode haver um qualificador, que costuma indicar a preparação prévia do alimento (por exemplo: “2 cebolas *grandes raladas*”, “300 g de tomates *sem pele e sem sementes picados*”, “1 pimentão vermelho *cortado em rodela*”). No inglês, os qualificadores antecedem o ingrediente, mas os que indicam seu pré-processamento ocorrem depois, geralmente precedidos por uma vírgula (p. e.: “3 large eggs, lightly beaten”, “1 small red onion, thinly sliced”, “1/2 stick cold butter cut into 4 pieces”).

Os participios não são comuns nas listas de ingredientes das receitas portuguesas (explicações equivalentes encontram-se no modo de fazer), mas o são na variante brasileira e em ambas as variantes do inglês (Tagnin e Teixeira, 2004b), sendo que no inglês americano é comum haver longas explicações de pré-preparo na lista de ingredientes (p. ex.: “2 cups torn Boston lettuce, washed well and spun dry”, “1 lb asparagus, washed, trimmed, and cut diagonally into 1-inch lengths”). As receitas

brasileiras costumam explicar procedimentos longos como esse no corpo do “modo de preparo”.

No caso de ingredientes pouco conhecidos, pode haver explicação ou indicação de onde consegui-los, em nota de rodapé ou entre parênteses, logo após o ingrediente: “1 colher de chá de mistura de garam masala (*pode ser adquirido em uma loja de comida indiana*)”; “1 colher de chá de ‘pimenta da Jamaica’ moída (*vide nota abaixo*)”; “100 gr de Snoubar (*pinholes*)”. Os parênteses também podem ser utilizados para dar informações adicionais sobre o ingrediente pedido: “6 a 8 bananas bem maduras (*nanica de preferência*)”; “1/4 do peito do frango (*peça para o açougueiro cortar em quatro mas manter o osso*)”.

Os temperos, acompanhamentos, ingredientes para polvilhar, guarnecer, pincelar, etc. são a exceção para as características mencionadas acima. Geralmente são os últimos itens da lista de ingredientes, não vêm acompanhados de uma quantificação precisa, podendo ocorrer mais de um ingrediente numa mesma linha. Exemplos: “sal e pimenta-do-reino a gosto”, “açúcar e canela a gosto”, “cebolinha picada fresca para guarnecer”, “Azeitonas pretas e ramos de ervas variadas para guarnecer”.

Quanto à ordem de aparição dos ingredientes, parece haver uma alternância de preferência. Alguns manuais de estilo sobre redação de receitas (Jacob, 2005: 177, por exemplo) sugerem que os ingredientes sejam dispostos na lista de acordo com a seqüência em que serão empregados no procedimento – se a preparação começar com um refogado, o óleo, a cebola e o alho serão os primeiros ingredientes da lista; se for uma de brócolis gratinado, tanto o brócolis quanto os ingredientes para o molho branco podem ser listados primeiro – vai depender de como será a descrição do preparo (“cozinhe o brócolis em água e sal... prepare o molho branco”, ou vice-versa).

Há casos, porém, em que a ordem seguida pode respeitar a importância do ingrediente para a elaboração do prato – em receitas de pães, por exemplo, a farinha costuma ser o primeiro ingrediente da lista, embora os ingredientes usados no pré-preparo do fermento sejam utilizados primeiro. Ostmann e Baker (2001: 8) sugerem que as carnes, peixes e aves, por seu valor comercial, devem sempre ser listados primeiro: **“Meat listed as first ingredient. Because meat (or fish, seafood, or poultry) is**

usually the main component and most expensive ingredient, it is an important factor in meal planning and grocery shopping”⁹⁶.

C. Modo de preparo

Esta parte da receita descreve os procedimentos a serem realizados com os ingredientes previamente listados de modo a obter o resultado esperado. O nome dado a esta parte da receita, que pode aparecer ou não na receita, varia não só de língua para língua, mas de cultura para cultura: as receitas brasileiras usam com mais frequência “modo de preparo” e “modo de fazer”; as portuguesas, “confeção”; as britânicas, “method” ou “instructions”; e as americanas “directions” (Teixeira 2004; Tagnin e Teixeira 2004a e 2004b).

Essas etapas podem ser escritas num parágrafo único, ou subdivididas em vários parágrafos, que podem ser enumerados, dispostos em forma de itens, ou não. As etapas devem seguir uma ordem lógica e cronológica. Todos os ingredientes listados devem ser mencionados no preparo e todos os ingredientes usados no preparo devem constar da lista de ingredientes.

Algumas características textuais do “modo de fazer” no português brasileiro incluem o uso preferencial do tempo verbal no modo imperativo e o fato de os ingredientes serem sempre precedidos de artigo definido (no inglês os artigos são, em geral, omitidos). Assim como no inglês, o objeto direto tende a ser apagado após as primeiras aparições, por exemplo: “Corte a carne em cubos médios. Tempere com sal e pimenta-do-reino. Frite em óleo quente até dourar. Reserve”. As sentenças começam geralmente com um verbo, ou, em alguns casos em que é indicado o uso de um recipiente específico, este costuma vir, preposicionado, no início da sentença. Exemplos:

Adicione as amêndoas e frite, mexendo sempre, até ficarem douradas. Tire do fogo e reserve. Numa panela grande, coloque o óleo Brejeiro e leve ao fogo brando para aquecer.⁹⁷

Numa tigela grande, amasse as bananas com o garfo. Acrescente a farinha, o sal, o açúcar, o coco, a água, a canela e as passas.

⁹⁶ *Carne como primeiro ingrediente da lista. Como a carne (ou peixe, frutos do mar ou frango) é, em geral, o componente principal e a o ingrediente mais caro, é um fator importante no planejamento da receita e na lista de supermercado.*

⁹⁷ <http://www.brejeiro.com.br/receitas/arroz.asp>

Misture bem e deixe a massa descansar por 30 minutos. Modele os biscoitos utilizando ½ colher (sopa) de massa.⁹⁸

O nível de detalhamento dos procedimentos vai variar conforme o público-alvo a que se destina a obra (que pode ser diferente para os textos de partida e de chegada numa tradução!). Jacob (2005) ilustra bem essa diferença de expertise ao comentar passagens de um livro escrito por um chef profissional para o público geral:

My favorite example of a chef trying to write a recipe for home cooks is an opening sentence a friend read to me: "Roast a duck as you would for any occasion." If you don't understand why this is funny, you won't be successful writing for a home cook.⁹⁹ (Jacob, 2005: 176)

Conhecer o grau de especialização do público-alvo a que se destina a receita (e sua tradução) – uma tarefa, a princípio, do editor, que deve passar as informações detalhadas para o tradutor para que este possa conformar o texto de chegada às características do usuário – é um dos aspectos mais importantes a se considerar na seção “modo de fazer”. Por envolver componentes culturais, históricos e até mesmo econômicos, a redação do “modo de fazer” das receitas tem gerado debates entre especialistas americanos:

At Kraft Foods, recipes never include words like "dredge" and "sauté." Betty Crocker recipes avoid "braise" and "truss." Land O'Lakes has all but banned "fold" and "cream" from its cooking instructions. And Pillsbury carefully sidesteps "simmer" and "sear." When the country's top food companies want to create recipes that millions of Americans will be able to understand, there seems to be one guiding principle: They need to be written for a nation of culinary illiterates.¹⁰⁰ (Sagon, 2006)

⁹⁸ <http://www.vitaminasecia.hpg.ig.com.br/culinariaintegral.htm>

⁹⁹ *Meu exemplo favorito de um chef tentando escrever uma receita para o público geral é uma sentença de abertura que uma amiga leu para mim: “Asse um pato como você o faria em qualquer ocasião”. Se você não consegue achar graça nisso, certamente suas receitas não terão sucesso junto a esse público.*

¹⁰⁰ *Na Kraft Foods, as receitas nunca incluem palavras como “dredge” [passar na farinha] e “sauté” [saltear]. As receitas da marca Betty Crocker evitam “braise” [brasear - refogar e cozinhar tampado e em pouco líquido] e “truss” [amarrar uma ave para assar]. A Land O'Lakes simplesmente baniu “fold” [juntar delicadamente] e “cream” [bater até obter um creme] das instruções culinárias de seus produtos, e a empresa Pillsbury evita cuidadosamente “simmer” [cozinhar em fogo muito brando] e “sear” [selar (uma carne)]. Quando as maiores empresas da indústria alimentícia do país tentam desenvolver receitas que milhões de norte-americanos serão capazes de entender, parece haver um único princípio norteador: elas precisam ser escritas para uma nação de analfabetos em culinária.*

People cook less and less. You don't want to intimidate them by using a term they don't understand. Doing so makes people feel stupid, then mad at themselves because they don't know the word.¹⁰¹ (Jacob, 2005: 181)

Além de evitar palavras "difíceis", Jacob (2005: 180) sugere que o uso de equipamentos caros e sofisticados, que o usuário pode não ter em casa (ela cita batedeiras de pé e processadores de alimentos), seja evitado. Entre os verbos listados pela autora e também por Ostmann e Baker (2001: 6) como de difícil compreensão estão: *blend* [misturar], *combine* [combinar], *cream* [bater (junto) até obter um creme], *cube* [cortar em cubos], *dice* [cortar em cubinhos], *fold* [juntar delicadamente] e *reduce* [(deixar cozinhar até) reduzir].

Instead of using these terms, describe the technique in simple words that are readily understandable to both novice and experienced cooks. For example, instead of saying "braise", say "simmer, covered". Many cooks frequently braise meat – they just do not use this term.¹⁰² (Ostmann e Baker, 2001: 6)

Por outro lado, Sagon (2006) cita Beth Wareham, editora da Charles Scribner¹⁰³, que ressalta o papel que as receitas podem ter na educação do usuário: "I tell them [writers and editors], 'Why should we dumb it [the recipe] down?' When you learn to drive, you learn terms like 'brake' and 'parallel park.' Why is it OK to be stupid when you cook?"¹⁰⁴. O tradutor tem que estar a par dessas discussões para tomar a decisão mais acertada na hora de traduzir.

Em receitas voltadas para um público mais especializado (vide item 2.3.4) o vocabulário pode ser mais técnico e específico, as medidas apresentadas apenas em quantidades precisas (que exigem o uso de balança de alta precisão, por exemplo) e pode-se lançar mão de equipamentos mais requintados.

¹⁰¹ As pessoas cozinham cada vez menos. Não se deve intimidá-las com o uso de um termo que não podem compreender. Isso faz com que se sintam ignorantes, e aborrecidas consigo mesmas por não conhecerem aquela palavra.

¹⁰² Em vez de usar esses termos, descreva a técnica em palavras simples, que possam ser compreendidas prontamente tanto pelo novato quanto pelo cozinheiro experiente. Por exemplo, em vez de dizer "braseado", diga "cozido em fogo baixo, com a panela tampada". Muitos cozinheiros usam a técnica de brasear em suas receitas – eles só não sabem que esse é o termo técnico.

¹⁰³ que publicou a edição comemorativa dos 75 anos do livro *Joy of Cooking*, um dos livros de receitas mais tradicionais e populares dos Estados Unidos.

¹⁰⁴ Eu digo a eles, 'Por que devemos idiotizá-las?'. Quando aprendemos a dirigir, aprendemos também o que é "ignição" e "baliza". Por que os cozinheiros devem manter sua ignorância.

If the recipe is directed to the professional chef, it should be short, clean and clear. It shouldn't insult the intelligence of the chef and should assume that the reader has an understanding of the fundamentals of cooking. It should make references to basic cooking principles rather than explain them.¹⁰⁵ (Ferdinand E. Metz, presidente do Culinary Institute of America, *apud* Ostmann e Baker, 2001: 136)

Vejamos, a seguir, alguns elementos opcionais das receitas culinárias.

2.3.3.2 Elementos opcionais

A. Comentários apreciativos

Esse item, apesar de não ser obrigatório, está presente na grande maioria das receitas. Assim como o título, tem a importante função de convencer o leitor a executar e/ou degustar o prato apresentado. Vem logo após o título, geralmente, ou em quadro destacado, ou ao final da receita. Pode trazer comentários apreciativos ou, no caso de ser uma receita assinada, relatar experiências pessoais do autor envolvendo a preparação e/ou a origem do prato, ou alguma explicação acerca de termos e/ou ingredientes envolvidos na preparação, notas históricas, etc. Exemplos:

O seu filho chegou com uma "turminha" para o lanche da tarde? Bateu uma fominha de última hora? Em qualquer uma destas situações esta pizza-sufle é uma boa pedida. Você pode inventar outros motivos para fazer esta receita é tão fácil [*sic*] e gostosa... Experimente.¹⁰⁶

Este é um dos mais famosos doces de ovos, ou melhor, de gemas, de origem portuguesa. Tornou-se muito popular nos últimos anos, sendo presença indispensável nas festas ou como sobremesa no almoço de Domingo. Esta receita é do começo do século e pertence ao acervo da família Cannabrava. Assim era preparado o quindim nas fazendas do interior de S. Paulo.¹⁰⁷

De acordo com Jacob (2005: 122), descrever um prato não é tarefa trivial. A autora propõe uma lista que subdivide os descritores em: i) **sabor e cheiro** – p. ex.:

¹⁰⁵ *Se a receita é voltada para o chef profissional, deve ser curta, simples e clara. Não deve insultar sua inteligência, mas assumir que esse leitor compreende os fundamentos da culinária. Deve, portanto, fazer referência aos princípios básicos da cozinha, em vez de explicá-los.*

¹⁰⁶ http://www1.uol.com.br/cybercook/colunas/cl_mo_pizzasufle.htm

¹⁰⁷ <http://www.novasociedade.com.br/cozinha/receitas/docinhos.html>

acrid (ácido, acre); peppery (apimentado, picante), buttery (amanteigado); ii) **textura** - p. ex.: velvety (aveludado), crisp (crocante, quebradiço); iii) **aparência** – p. ex.: caramelized (caramelizado/a), melted (derretido), sprinkled (polvilhado, sapicado); iv) **som** – p. ex.: bubbling (borbulhante), popping (que explode(m)), crackling (crocante); v) **outros**: satisfying (que satisfaz), alluring (tentador/a).

A autora observa ainda que a maioria das introduções em receitas atuais traz notas de cunho pessoal; ela sugere que as notas de fundo cultural, histórico, instrucional (detalhando etapas do preparo ou ingredientes desconhecidos) e sensoriais sejam mais exploradas (Jacob, 2005: 173).

Ao traduzir essa seção das receitas, o tradutor deve ter em mente, sempre, o objetivo desse subtipo textual: informar o leitor sobre o que ele deve esperar do prato, atraindo sua atenção e tornando a receita convidativa, interessante. Muitas vezes isso exigirá mudanças no texto. Exemplos¹⁰⁸:

- “This is a deliciously indulgent, creamy sauce” → “Este molho é um deleite para os sentidos com sua textura cremosa”;
- “This spicy sauce, enlivened with chilli and vodca, might be the answer when you fancy a quick easy lunch or supper to perk up jaded tastebuds” → “Este molho estimulante, com um toque de pimenta fresca e vodka, pode ser a resposta certa para quando você quer um prato fácil de fazer mas que ‘desperte’ as papilas gustativas”;
- “This is basically a macaroni cheese for meat lovers” → “Esta é uma variação da tradicional de receita de *macaroni and cheese*^{NT}. (N.T.: Também conhecido por *macaroni cheese* no Reino Unido e *mac’n’cheese* no Canadá e nos E.U.A., consiste em combinar massa curta seca com um molho cremoso de queijo – geralmente o cheddar. A receita tradicional vai ao forno para gratinar.);

B. Dicas diversas

As receitas podem conter vários tipos de dicas. Elas podem aparecer em partes diferentes da receita, entre parênteses, em forma de nota de rodapé / nota de fim, ou até mesmo no corpo do texto (Comentários, Lista de ingredientes ou Modo de preparo):

¹⁰⁸ retirados de livros de receita que já traduzi.

- notas sobre os ingredientes (onde encontrá-los, como escolhê-los, como estocá-los, etc.) – p. ex. “1 pé pequeno de salsão (escolha os mais novos)”, “Ferva a água para o arroz com as cabeças dos camarões previamente limpas sem os olhos e sem a sujeira esverdeada de dentro (limpe inserindo o dedo indicador, sob água corrente)”;
- dicas de substituição de ingredientes específicos, ou de variação da receita como um todo – p. ex.: “50g de *pancetta* (se não encontrar, use bacon)”;
- dicas sobre técnicas específicas do preparo – p. ex.: “Quando for assar a massa da torta, ainda sem recheio, fure-a com um garfo, para que não cresça.”, “leve-a à geladeira por 15 minutos para ficar mais fácil de trabalhar”;
- alertas sobre como evitar problemas comuns e, caso ocorram, como remediá-los – p. ex.: “Cuidado para não bater demais, pois o creme chantilly pode virar manteiga.”, “Se a maionese desandar, junte uma colher de sopa de água fria e continue batendo”;
- dicas de apresentação ou acompanhamento – p. ex.: “Sirva com arroz branco”, “Sirva gelado, polvilhado com açúcar e canela”.

C. Recursos icônicos

A maioria das receitas em livros e *sites* de culinária apresenta uma foto do prato final e, quando voltadas para o público mais inexperiente, fotos das etapas de preparo (vide item 2.3.2). Essas fotos podem aparecer lado a lado com a receita ou, em edições mais baratas, concentradas em um ou vários miolos de fotos coloridas, que são intercalados entre as receitas, impressas num papel mais barato – uma foto culinária colorida com qualidade profissional pode custar entre U\$700 e U\$1.000, nos Estados Unidos (Jacob, 2005: 161).

Para o tradutor, o que importa é observar as possíveis discrepâncias entre a foto e o que está sendo pedido na receita. A troca de ingredientes, por exemplo, pode ser problemática, já que não existe a possibilidade de mudar a foto. Na tradução a seguir, por exemplo, a receita pede “castanha de caju”, mas a foto mostra o “pinhole”:

Figura 9: exemplo de inconsistência entre receita e foto

Ao traduzir um livro sobre Massas recentemente me deparei com uma receita de *Minestrone* que pedia “cavolo nero” e outra de *Tagliatelle with spicy pea fritters* que pedia “split yellow peas” – ambas com fotos do prato final e das etapas do preparo.

No primeiro caso, deixei o nome da verdura original, com uma nota de tradução explicativa, e acrescentei a opção de substituição indicada pelas obras de referência consultadas: “espinafre”. No segundo, a receita sofreu várias alterações – no título, nos comentários apreciativos, na lista de ingredientes e no modo de fazer – pois o editor optou por trocar as *yellow peas* por “grão-de-bico” (felizmente, uma única foto dos passos da receita que mostra as *yellow peas* está desfocada na região em que elas estão, sendo impossível identificar o ingrediente – vide Fig. 10).

<p>Tagliatelle with spicy pea fritters <i>These little split pea fritters are quite hot and spicy, so halve the spices for a milder flavour. They can be made in advance and kept in the refrigerator.</i></p> <p>PREPARATION TIME: 30 minutes, plus soaking COOKING TIME: 40 minutes SERVES: 4</p> <p>300 g (10 oz) split yellow peas 1 onion, roughly chopped 25 g (1 oz) breadcrumbs 2 garlic cloves, chopped 1 tablespoon crushed cumin seeds ¼ teaspoon dried chilli flakes several sprigs of mint 1 egg 8 tablespoons lemon-infused olive oil 25 g (1 oz) butter 200 g (7 oz) dried tagliatelle 200 g (7 oz) roasted red peppers, thinly sliced 4 tablespoons chopped coriander salt and pepper</p> <p>1 Put the peas in a large bowl, cover with cold water and leave to soak overnight. Drain and put in a saucepan. Cover with cold water, bring to the boil and cook for about 25 minutes until tender. Drain.</p>	<p>Talharim com bolinhas picantes <i>Estas bolinhas de grão-de-bico são bem temperadas e picantes – reduza os temperos pela metade para obter um sabor mais suave. Elas podem ser feitas com antecedência e guardadas na geladeira.</i></p> <p>TEMPO DE PREPARO: 30 minutos, mais o tempo para deixar de molho TEMPO DE COZIMENTO: 40 minutos RENDIMENTO: 4 porções</p> <p>300 g de grão-de-bico 1 cebola grosseiramente picada 25 g (1/2 xícara) de miolo de pão esmigalhado 2 dentes de alho picados 1 colher (sopa) de sementes de cominho trituradas ¼ colher (chá) de pimenta calabresa em flocos vários raminhos de menta (ou hortelã) 1 ovo 8 colheres (sopa) de azeite aromatizado com limão siciliano 25 g de manteiga 200 g de talharim seco 200 g de pimentão vermelho assado, sem pele e sem sementes, cortado em tiras finas 4 colheres (sopa) de coentro picado sal e pimenta-do-reino</p> <p>1 Deixe o grão-de-bico de molho de um dia para outro. Escorra e coloque em uma panela. Cubra com água fria, ferva e cozinhe por cerca de 45 minutos, ou até ficar macio. Escorra.</p>
---	--

Figura 10: Exemplo de alteração de ingrediente compatível com a foto¹⁰⁹

Quando o ingrediente não está disponível (ainda!) para venda no país do texto de chegada, ou é difícil de encontrar, dependendo do público-alvo e do tipo de publicação (algumas não admitiriam uma nota de tradução, por exemplo), é mais vantajoso manter o ingrediente original, acrescido de nota de rodapé, de tradução ou de explicação entre parênteses, e oferecer uma opção de substituição – que pode ser pesquisada em obras de referência ou por meio de uma consulta a especialista. Assim, a obra será mais fiel ao original e se manterá atual caso o ingrediente em questão venha a ser comercializado em larga escala no futuro.

D. Rendimento

Informa ao leitor o número de pessoas que o prato final serve (p. ex.: “Serve 6 pessoas”, “Dá para 4 pessoas”, “Rendimento: 8 a 10 porções”), ou quantas porções de um determinado tamanho o prato rende (p. ex.: “Rende 1 bolo de 20cm de diâmetro”, “Rendimento: 50 brigadeiros pequenos”). Não costuma causar problemas para a tradução, a não ser nos casos em que o autor da receita comete erros. Por exemplo:

¹⁰⁹ retirado de Farrow, J. (2007). *Massas: 80 das melhores receitas*. Tradução de Elisa D. Teixeira. São Paulo: Manole. A versão em português não contém a foto pois não é a versão publicada, mas a prova que a editora me enviou para revisão final da tradução.

ao traduzir uma receita de sanduíche do livro *Take charge of the change*¹¹⁰, “Eggless egg salad sandwich”, percebi que todos os ingredientes pedidos eram para 5 porções (5 folhas de alface, 10 rodellas de tomate e 15 de pepino), com exceção do pão (que pedia 4 fatias e, ao final da receita, havia: “makes 4 sandwiches”) – mudei o número de fatias de pão para 10 nos ingredientes e o rendimento para 5 porções.

E. Informações nutricionais

A melhora no padrão de vida das camadas menos favorecidas das sociedades desenvolvidas, associada à popularização das redes de *fast food* e de alimentos industrializados, causou uma epidemia de obesidade e má nutrição nas últimas duas ou três décadas. Nos Estados Unidos a situação é mais alarmante, mas já se vêem sinais desse problema também no Brasil.

Em resposta a isso, e à crescente valorização do indivíduo e da aparência física sugestionada pela mídia, as dietas se tornaram cada vez mais populares, sejam as de emagrecimento, sejam as dedicadas a condições médicas específicas, como hipercolesterolemia, diabetes, pressão arterial elevada, etc. Toda dieta consiste no controle do que é ingerido, e as tabelas de informação nutricional, que também aparecem em rótulos de produtos e, atualmente, até mesmo em cardápios de alguns restaurantes, são essenciais para monitorar a qualidade e a quantidade de alimentos consumidos pelo indivíduo.

As receitas que apresentam uma tabela de informações nutricionais (em que pode constar apenas o número de calorias oferecido por cada porção, mas em geral são dadas as quantidades de fibras, gorduras, sódio, vitaminas, etc.) apresentam um complicador extra para a tradução. Caso seja necessário alterar algum ingrediente, essa tabela deve ser reformulada – ou, o ingrediente escolhido para a substituição deve ter fatores nutricionais muito semelhantes aos do ingrediente original. De qualquer modo, se as informações nutricionais forem de suma importância para a obra (um livro de dieta, por exemplo), o tradutor e/ou editora devem contratar um profissional da área de nutrição para que os valores sejam recalculados.

¹¹⁰ SMITH, P. (2005). *Mulheres de bem com a vida*. Tradução: Lenita M. R. Esteves e Elisa D. Teixeira. São Paulo: Editora Vida.

F. Categorizações

Alguns livros e vários *sítes* de receitas usam critérios diversos para classificar as receitas. Esses critérios facilitam a localização numa busca – por exemplo: quando o usuário tem em mente um resultado (bolo), um ingrediente (berinjela), um tipo de processo (refogado), um estágio da refeição (entrada), etc. em mente e precisa encontrar receitas correspondentes.

Algumas das categorias mais usadas para classificar receitas são:

- **tipo de prato:** a primeira grande subdivisão é doce / salgado, mas há várias outras “oposições” usadas para classificar os pratos, como: culinária nacional / regional; culinária nacional / internacional; dia-a-dia / ocasiões especiais; culinária caseira / requintada; pratos quentes / frios; culinária light / diet. Algumas categorias comumente usadas para classificar os pratos por seu tipo: bolos, molhos, churrasco, sanduíches, salgadinhos, saladas, massas, pães, tortas, sopas, sorvetes, compotas, doces, guarnições, pizzas, pavês, pudins, etc.
- **estágios da refeição:** aperitivo, entrada, prato principal, acompanhamento, sobremesa, lanche (da tarde), café-da-manhã, etc.
- **tipo de ingrediente:** critério mais aplicado aos pratos salgados. Exemplos: Carnes, Ovos, Massas, Legumes e Verduras, Peixes e frutos do mar, Aves, Grãos e cereais, etc., mas também: receitas com chocolate, sem glúten, utilizando produtos de marcas conhecidas, etc.
- **grau de dificuldade:** baseado no nível de dificuldade das técnicas empregadas e no número de etapas necessárias para concluir o prato.
- **tempo de preparo:** geralmente inclui o tempo de pré-preparo (p. ex.: “deixe o feijão de molho de um dia para o outro”) e pós-preparo (p. ex.: “leve ao congelador por no mínimo 4 horas antes de servir”)
- **custo:** calculado com base no preço médio dos ingredientes utilizados

G. Outras informações

Algumas receitas apresentam, logo após ou imediatamente antes da lista de ingredientes, uma lista dos utensílios e/ou equipamentos que serão necessários para a execução da receita. Exemplo:

BOLO COM CALDA DE RUM¹¹¹

♦ Dificuldade: média

♦ 12 fatias

♦ Calorias: 322 por fatia

♦ Exige preparo antecipado

Ingredientes:*Massa*

2 xícaras de farinha de trigo
 1/2 xícara de água morna
 1/4 xícara de manteiga amolecida
 2 colheres (sopa) de açúcar
 1/2 colher (chá) de sal
 3 ovos

1 tablete de fermento biológico fresco (15 g)

Calda

1 1/2 xícara de açúcar
 1 xícara de suco de laranja
 3/4 xícara de água
 3 1/2 colheres (sopa) de rum

→ Acessório: forma de buraco no meio de 23 cm de diâmetro e capacidade para 3 litros

Preparo:*Massa*

1. Unte a forma com manteiga. Misture o fermento com a água. Adicione o açúcar, o sal, os ovos, a manteiga e uma xícara e meia da farinha. Bata na batedeira por 2 minutos ou até a massa ficar lisa. Junte a farinha restante e misture até ficar homogêneo.

2. Coloque a massa na forma e deixe descansar até dobrar de volume. Aqueça o forno em temperatura alta. Asse por 15 minutos ou até dourar ligeiramente. Desenforme ainda quente e coloque num prato.

Calda

Ferva o açúcar com a água em fogo alto por 5 minutos. Junte o suco de laranja e o rum. Retire do fogo. Pincele o babá com a calda quente. Sirva ao esfriar.

2.3.4 Receitas quanto ao grau de especialização

Baccin (2003) propõe uma divisão da “terminologia alimentar” em três subáreas:

i) nutrição e dietologia; ii) gastronomia e arte culinária; iii) cozinha do dia-a-dia. Para cada subárea, a autora fornece definições¹¹², conceitos subjacentes, os meios sociais onde é utilizada, o público usuário, o tipo de documento em que comparecem, além de exemplos de verbos e substantivos típicos (vide Tab. 1)

¹¹¹ retirado do CD *Grandes Receitas Cláudia Cozinha* com 6 edições especiais da Revista Cláudia Cozinha, distribuído pela editora Abril, sem data de publicação.

¹¹² retiradas da versão eletrônica do Dicionário Aurélio século XXI, que omitimos na reprodução de seu quadro-resumo na Fig. 11.

área	terminologia alimentar		
subáreas	nutrição e dietologia (T1)	gastronomia e arte culinária (T2)	cozinha do dia-a-dia (T3)
conceitos subjacentes	sustento e saúde	arte e prazer	sustento, arte e prazer
onde é encontrada / utilizada	faculdades de nutrição e engenharia de alimentos; química; biologia dos alimentos	cursos profissionalizantes de gastronomia e hotelaria	cursos de culinária para amadores; no ambiente doméstico; no dia-a-dia
público usuário	profissionais de nutrição e alimentos; médicos	<i>chefs</i> ; gourmets; gourmands	todos os falantes em situações específicas
documentos	livros científicos; teses; artigos e revistas especializadas em alimentação; dietética e nutrição	livros de receitas, revistas, jornais	livros de receitas; cadernos pessoais de receitas; tradição oral; revistas; jornais.
exemplos de verbos	alimentar; nutrir	degustar	comer
exemplos de substantivos	alimento; dieta; regime alimentar	iguarias; acepipes; pratos; alimento	comida; pratos

Tabela 1: terminologia alimentar de acordo com Baccin (2003)

A divisão proposta por Baccin mostra-se pertinente em vários pontos. A terminologia alimentar, enquanto temática, gravita entre três domínios principais: dietética, arte e sustento. Se de um lado pode-se identificar o nutricionista, o profissional que trabalha no setor de restaurantes e a cozinheira do dia-a-dia como usuários preferenciais dessas terminologias, também é possível divisar os respectivos ambientes em que tais terminologias circulam – o acadêmico, o profissional e o doméstico. No entanto, com o processo acelerado de globalização e o *boom* da Culinária na mídia, a divisão que a autora faz da terminologia alimentar no tocante aos graus de especialização parece estar se tornando cada vez menos pertinente, especialmente no que concerne à distinção entre culinária sustento / culinária arte. Como mostram outras passagens de seu texto, essa distinção, de fato, não pode ser facilmente estabelecida:

O público usuário de T2 é formado por *chefs*, cozinheiros profissionais, críticos gastronômicos, etc., especialistas em gastronomia e **especialistas em T2**. Esse público é formado também por **usuários não-especialistas em T2**. São falantes da língua geral que se dedicam à gastronomia como *hobby*. (Baccin, 2003, grifos da autora).

A subárea 1 é a mais específica e tem um público de usuários numericamente menor; é o mais especializado, pois os usuários são todos especialistas. A subárea 2 tem usuários especialistas (profissionais da área) e os usuários não especialistas (amadores e curiosos). A subárea 3 não possui especialistas entre os usuários. Seu público é constituído pela totalidade da população de falantes, já que todos se servem de T3 em determinadas circunstâncias, não há, portanto, especialistas em T3. (*id. ibid.*)

Se, por um lado, concordo com Baccin (2003) quando diz que o público usuário da terminologia da gastronomia e arte culinária (T2) compreende não apenas especialistas e profissionais da área, mas não-especialistas (“amadores e curiosos”, nas palavras da autora), discordo que em T3 (terminologia alimentar da cozinha) não haja especialistas, pois seria o mesmo que afirmar que as cozinheiras de longa data estejam num nível de especialização inferior ao dos “curiosos e amadores”, ou mesmo do aprendiz de *chef* que acaba de iniciar seu curso profissionalizante. Da mesma forma, não se pode dizer que um livro de dieta, seja ele voltado para o *chef* profissional, seja para o público-geral não-especializado, não contenha vocabulário especializado da terminologia alimentar da nutrição e dietologia (T1).

Jean-François Revel, em seu livro *Um Banquete de Palavras*, ao comentar sobre a origem da cozinha e seus pratos diz:

A cozinha procede de duas fontes: uma popular e a outra erudita, esta necessariamente situada nas classes ricas de todas as épocas. Ao longo da história, há uma cozinha camponesa (ou marinheira) e uma cozinha de corte; uma cozinha plebéia, familiar, executada pela mãe de família – ou a humilde cozinheira doméstica – e uma cozinha de profissionais que só os *chefs* fanáticos, inteiramente dedicados à pesquisa, têm tempo e ciência para executar. (Revel, 1996: 38)

Se essa distinção for aplicada às subáreas propostas por Baccin, veremos que para cada uma delas há vários graus de especialização possíveis. A subárea de “nutrição e dietologia” (T1), por exemplo, vai estar presente não só nos compêndios médicos e nos livros de receita da alta cozinha, mas também nas dicas de alimentação saudável veiculadas pelo jornalzinho de ofertas do supermercado. Dessa perspectiva, a subárea “nutrição e dietologia” deixa de ter interesse para este estudo enquanto

categoria isolada - interessará na medida em que comparecer tanto em textos culinários voltados para o público-geral quanto nos voltados aos especialistas.

Também do ponto de vista dos níveis de especialização, veremos que tanto na subárea da “gastronomia e arte culinária” quanto na de “cozinha” haverá especialistas e não-especialistas – se há o especialista “institucionalizado”, da esfera pública (o *chef*, que teve educação formal e/ou atua profissionalmente no setor) no primeiro caso, há também o especialista da esfera doméstica (a cozinheira, a quituteira, etc.) no segundo. No outro extremo da não-especialização, há o “*gourmand*” e o “cozinheiro de fim-de-semana” como representantes do não-especialista interessado em gastronomia e arte culinária, e o público em geral (recém-casada, adolescente que vai morar sozinho e começa a cozinhar, etc.) como não-especialistas interessados em “cozinha”.

Assim, por ter a tradução como foco do meu trabalho, proponho uma redistribuição das subáreas apresentadas por Baccin de um ponto de vista mais textual, uma vez que isso afetará diretamente o tipo de terminologia veiculado. Em vez de categorias isoladas, teríamos um continuum representando, no eixo x, os graus de especialização do público-alvo a que os textos culinários se destinam, e no eixo y, o grau de penetração nas esferas pública / profissional e doméstica / popular dessa terminologia.

Figura 11: níveis de especialização do ponto de vista textual

A Culinária publicada atualmente se situaria entre esses extremos, seja no que diz respeito ao grau de especialização, seja do ponto de vista das esferas de circulação. Retomando Revel (1996: 38), isto não quer dizer que os "*chefs* fanáticos", e as mães de família que cozinham o "trivial" todos os dias não mais existam, mas há uma Culinária que está entre esses dois extremos e que, associando as técnicas daquela ao saber ancestral desta, está criando uma cozinha que poderia ser chamada de 'globalizada' – alvo de um grande volume de traduções e versões, sejam os textos

receitas, rótulos de produtos importados, *sítes*, revistas especializadas, artigos em jornais, programas televisivos, etc.

Aqueles singelos caderninhos de receitas, cevados durante décadas pelas mulheres da casa – as cozinheiras de forno e fogão, que os preenchiam com receitas maravilhosas e caligrafia feminina –, são hoje obras de referência. Mereceriam destaque em museus, com todo o louvor. Agora o que se tem são edições sofisticadas, tecnicamente bem-feitas e consumidas majoritariamente por cozinheiros eventuais – gente de pouco trato com as panelas, mas imensa curiosidade sobre aromas e sabores. (...) “Os livros de gastronomia são disputados pelo público masculino, que encara a culinária como hobby”, garante Alexandre Dórea Ribeiro, da DBA Melhoramentos. (Barone, 1999)

Essa “cozinha globalizada” – a despeito do desprestígio histórico da culinária, por ser atividade originalmente associada ao mundo feminino e a uma *praxis* – vem se investindo de uma carga ideológica que a identifica como símbolo de status socioeconômico e refinamento cultural / intelectual. Para muitos autores, isso se deve ao interesse crescente do público masculino pelo tema.

2.3.5 Receitas e estilo

Em língua inglesa, há uma série de livros sobre como escrever todo tipo de texto nas áreas de culinária e gastronomia, e até mesmo manuais de estilo voltados só para a redação de receitas e livros de receitas (p. ex.: Ostmann e Baker, 2001; Whitman e Simon, 1993; Jacob, 2005; Allen, 1999; Hullah, 1984; Comfort, 1997). Muitas empresas e editoras especializadas têm seus manuais de estilo próprios: *The Betty Crocker Stylebook*, da General Mills Inc.; *Cookbook Guidelines*, da Simon & Schuster; *Kraft General Foods Recipe Style Manual*, da Kraft General Foods; *Recipe style*, do jornal Chicago Tribune, etc. (Ostmann e Baker, 2001: 297). Ou seja, o tradutor que for verter receitas para o inglês conta com uma excelente bibliografia de apoio para efetuar as mudanças necessárias de estilo, seja no nível micro-, seja no nível macrotextual.

Em língua portuguesa, não tenho conhecimento de publicações desse tipo. O livro que mais se aproxima dessa proposta é *Berinjela se escreve com j*, de Josimar Melo (1999), que trata da grafia correta de nomes de pratos e ingredientes em cardápios de restaurantes. Azenha Jr. (1999) observa essa disparidade de recursos também no par de línguas português e alemão, que faz com que os textos técnicos

escritos nesta língua, em particular os de instrução, sejam mais bem escritos. Comenta: “talvez pelo fato de as linguagens técnicas já virem sendo descritas e discutidas na Alemanha há algum tempo” (p. 123). Entre os problemas identificados nos textos do mesmo tipo em português, o autor cita a falta de elementos coesivos (na opinião dele, devido a “uma noção de bom estilo – segundo a qual não se deve repetir um mesmo elemento – transplantada de outras tipologias de texto” (p. 124)) e a profusão de erros de ortografia, acentuação, regência, etc., “que apontam para um processo de revisão deficitário e reforçam a hipótese de que o texto final não passa pelo crivo de leituras críticas mais apuradas” (p. 124).

A carência de materiais de referência que possam nortear o tradutor e o revisor na redação de textos culinários em português brasileiro, aliada à “noção de bom estilo” mencionada por Azenha Jr., advinda de manuais de estilo escritos para textos de uma tipologia totalmente diferente, podem conduzir, inclusive, ao outro extremo da situação – o erro causado pela hipercorreção, como mostram os exemplos¹¹³ da tabela 1:

texto entregue pelo tradutor	texto revisado pela editora (e publicado)
Cozinhe o linguine em bastante água fervente, temperada com sal, até que esteja <i>al dente</i> (cerca de 12 minutos). Escorra.	Cozinhe o linguine em bastante água fervente, temperada com sal até que esteja <i>al dente</i> (cerca de 12 minutos). Escorra- o .
Junte o frango e refogue, mexendo de vez em quando, até que esteja quase cozido (3 a 4 minutos). Retire da panela e reserve. Junte outra colher de sopa do óleo à panela. Adicione o tofu e refogue, mexendo, por 2 minutos. Reserve, junto com o frango.	Junte o frango e refogue, mexendo de vez em quando, até que esteja quase cozido (3 a 4 minutos). Retire- o da panela e reserve- o . Junte outra colher de sopa do óleo à panela. Adicione o tofu e refogue- o , mexendo, por 2 minutos. Reserve- o , junto com o frango.
Cubra o prato com filme plástico – faça alguns furos para que o vapor possa sair. Leve ao microondas, em potência alta, por cerca de 5 minutos.	Cubra <u>o prato</u> com filme plástico – faça alguns furos para que o vapor possa sair. Leve a forma ao microondas, em potência alta, por cerca de 5 minutos.
Caso percam sua crocância, aqueça-as no forno a 180° por 5 minutos antes de servir.	Caso deixem de estar crocantes , aqueça-as no forno a 180° por 5 minutos antes de servir.
Leve ao fogo, numa panela de fundo grosso, o azeite de oliva e o alho. Frite por cerca de 1 minuto. Junte os temperos e o espinafre.	Leve ao fogo, numa panela de fundo grosso, <u>o azeite de oliva e o alho</u> . Frite esses ingredientes por cerca de 1 minuto. Junte a eles os temperos e o espinafre.

Tabela 2: exemplos de hipercorreção na revisão de receitas por editoras

Outra questão relacionada ao estilo na redação de receitas envolve a autoria. As receitas no geral são textos formulaicos de grande circulação em todas as camadas da sociedade e sem autoria definida – são de domínio público, por assim dizer¹¹⁴ (vide por

¹¹³ retirados de livros de receita que já traduzi.

¹¹⁴ tanto a lei dos direitos autorais brasileira quanto a americana não protegem materiais pré-existentes ou processos, instruções ou métodos (a maioria das receitas poderia ser incluída nessas categorias)

exemplo o grande número de obras sem indicação de autor ou data de publicação no levantamento feito por Gomes e Barbosa (2004: 5) dos livros sobre culinária publicados no último século no Brasil). Há inúmeras versões das receitas mais populares, com ligeiras variações: pão de queijo, lasanha, feijoada, quindim, pudim de leite, bolo de chocolate, etc. O “autor”, nas receitas tradicionais dos cadernos de nossas mães e avós, é mais bem referido como “autoridade”, ou seja, “quem passou a receita” – a “brevidade” da Comadre Nilza, a “rosca” da Tia Celina, o “lombinho” da D. Teresa, etc. Essa tendência de atribuir um “dono / dona” para a receita começou a aparecer no mercado editorial brasileiro na década de 1940 – *Cozinhando com D. Benta* (1940), *Receitas de Yayá Ribeiro* (1966), *300 receitas de dona Mimi* (1970), *Receitas selecionadas de Ofélia* (1979), etc.

As receitas culinárias, por constituírem uma tradição não-autoral, têm na variação terminológica um fator identitário – ou seja, a variação terminológica vai, muitas vezes, permitir a associação de uma receita a um determinado período histórico, um público-alvo específico, um veículo de comunicação em particular, etc. Para explorar esse tipo textual num mercado autoral é preciso dar uma contribuição mais pessoal à receita, particularizando-a – seja na maneira de reunir, num só volume, certas preparações em detrimento de outras, seja oferecendo uma dica de preparo, de escolha de ingredientes, de apresentação, etc. Nas palavras de Jacob (2005: 166): “A well-written recipe is poetry, like beautiful writing. (...) A badly written recipe is a disappointment.”¹¹⁵.

A apresentadora de TV Ana Maria Braga, por exemplo, há alguns anos, optou por deixar uma marca ainda mais pessoal em suas receitas. Em vários de seus livros mais antigos – por exemplo, *As receitas preferidas de Ana Maria Braga* (Círculo do livro, 1997) –, além de inserir diálogos com o mascote do programa ao pé da página, a apresentadora colocou todos os verbos do modo de preparo na primeira pessoa do singular, no presente do indicativo (vide Fig. 12). Como pode ser visto na versão atual do site do programa *Mais Você*¹¹⁶, essa estratégia não persistiu por muito tempo.

¹¹⁵ Uma receita bem escrita é poesia, como um belo texto. (...) Uma receita mal escrita é uma decepção.

¹¹⁶ <http://receitas.maisvoce.globo.com/>

Macarrão do campo

Ingredientes	Preparação
<p>1 pacote de macarrão tipo linguine ou espaguete</p> <p>2 colheres (sopa) de óleo</p> <p>1 xícara (chá) de bacon (toucinho defumado) picado</p> <p>2 xícaras (chá) de cebola picada</p> <p>4 dentes de alho bem picados</p> <p>1 xícara (chá) de salsa bem picada</p> <p>sal e pimenta-do-reino a gosto</p> <p>1 kg de batata cozida al dente em água temperada com sal, descascada e cortada em cubinhos</p>	<ul style="list-style-type: none"> • Coloco bastante água numa panela grande, tempero com sal a gosto, levo ao fogo alto e deixo ferver. Junto o macarrão de uma só vez, deixo abrir fervura novamente e cozinho até o macarrão ficar <i>al dente</i>. • Enquanto isso, preparo o molho: coloco o óleo e o bacon numa panela, levo ao fogo médio e frito até o bacon ficar bem dourado e crocante. Então, acrescento a cebola e o alho e refogo, mexendo de vez em quando, até a cebola ficar macia. Adiciono a salsa e misturo bem. Tempero com sal e pimenta a gosto e tiro do fogo. • Tiro também o macarrão e escorro bem. Coloco num prato de servir, acrescento os cubinhos de batata e o refogado de salsa, misturo cuidadosamente e levo à mesa.
<p>RENDIMENTO: 6 porções</p>

Louro José: – Quando eu arrumar uma noiva, esta receita vai para o casamento!

Ana Maria: – *Combinado, Louro José! Eu faço uma macarronada maravilhosa para você!*

Figura 12: exemplo de receita com verbo na 1a. pessoa do singular

Ostmann e Baker ressaltam a importância da receita bem escrita, mas enfatizam que a parte mais adequada da receita para o autor expressar seu estilo e interagir com o leitor é nos comentários apreciativos:

Many people read recipes and cookbook for enjoyment, so writing style is almost as important as cooking reliability. Although certain basic information should be included in every recipe, the way you word this information can reflect your personal style and entice the reader. The headnotes on recipes provide an especially good place for the author's voice to shine through.¹¹⁷ (Ostmann e Baker, 2001: 11)

¹¹⁷ Muitas pessoas lêem receitas e livros culinários por prazer. Portanto, o estilo nessas obras é quase tão importante quanto a precisão. Apesar de certas informações básicas serem obrigatórias em qualquer receita, a maneira como são ditas pode refletir o estilo pessoal do autor e prenderem o leitor. Os comentários que vêm logo após o título nas receitas são um ótimo lugar para o autor expressar seu estilo.

Enquanto não houver manuais de estilo voltados para textos das áreas de Culinária e Gastronomia em português brasileiro, pode-se lançar mão de corpora, *sites* confiáveis da Internet e/ou livros de editoras consagradas para tentar minimizar essa lacuna. Como já foi mencionado no item 2.1.2, cabe ao tradutor: primeiramente, exigir a participação no processo de revisão e, tendo esse direito assegurado, apresentar materiais convincentes aos revisores e editores para assegurar que suas escolhas prevaleçam.

2.4 Tradução culinária no par de línguas inglês – português

Ainda que bons livros de culinária sejam escritos em várias línguas, e que o francês seja a língua da gastronomia por excelência, a maioria dos livros traduzidos da área publicados no Brasil atualmente parece entrar no país via língua inglesa. Afinal, 25% de toda a bibliografia sobre culinária publicada no mundo nos últimos 10 anos foram escritos nessa língua – o espanhol ocupa o segundo lugar da lista, com cerca de 16,5%, o francês é apenas o terceiro, com 12.7% (Cointreau, 2005), e o português sequer figura individualmente.

Alguns problemas impostos pela tradução de receitas são os mesmos enfrentados na tradução de todo tipo de texto, outros são específicos desse gênero, como veremos no item 2.4.1 a seguir. Por serem textos fortemente marcados pela cultura, conforme o par de línguas trabalhado, as receitas vão impor certas dificuldades ao tradutor. No item 2.4.2, apresento algumas ferramentas de apoio à atividade tradutória, dando ênfase especial ao par de línguas inglês-português. Por fim, faço um apanhado das necessidades do profissional da área no item 2.4.3 e a forma como pretendo abordá-las em meu modelo de verbete eletrônico voltado para o tradutor.

2.4.1 Problemas de não-equivalência e equivalência parcial em Culinária

Em Teixeira (2004), no capítulo em que trato das especificidades da tradução de receitas culinárias, subdivido os principais problemas de não-equivalência e equivalência parcial encontrados na tradução de um livro de receitas dietéticas inglês-português em categorias abrangentes: i) ingredientes difíceis de encontrar no Brasil; ii) ingredientes não disponíveis no Brasil; iii) erros no original e iv) diferenças culturais.

Com o intuito de aprofundar um pouco mais esse estudo, de modo que possam ser previstos problemas de equivalência parcial e não-equivalência em Culinária, que serão possivelmente encontrados quando forem estabelecidas as equivalências entre os termos pesquisados nos *corpora* de estudo, retomemos os modelos de Barbosa (1990: 93) e Schmitt (1989 apud Azenha Jr., 1999: 79) apresentados no item 1.2.5.2 do Capítulo I.

Primeiramente, vejamos se é possível aplicar à Culinária todas as categorias de “incongruências conceptuais interculturalmente condicionantes” identificadas por Schmitt:

A. Diferenças nas hierarquias conceptuais

→ Em língua inglesa, os *muffins* (bolinhos assados em fôrmas individuais semelhantes às de empada) são classificados como pães, ao passo que no português brasileiro entram na categoria bolos; categorias semânticas como “banana” (nanica, ouro, prata, maçã, são tomé, d’água, etc.) e “manga” (espada, tommy, coquinho, coração de boi, rosa, bourbon, etc.), no inglês, certamente têm menos subtipos que em português – o inverso é verdadeiro para categorias como “frutas silvestres” (*blueberry, blackcurrant, cranberry, gooseberry, boysenberry*), e “cogumelos” (*chestnut, roman brown, chanterelle, straw, portabella, oyster*, etc.);

B. Diferenças na construção de produtos condicionadas por normas e leis específicas de cada cultura

→ Na Culinária, não se trata tanto de leis, ainda que os livros de receita geralmente tragam textos padronizados (vide item E desta lista) alertando sobre o uso de ingredientes que podem causar alergias, doenças, etc. No entanto, um problema similar ocorre na tradução dos tamanhos padrão para fôrmas e panelas, que variam de uma cultura a outra, nos diferentes cortes de carne bovina, suína e caprina / ovina, ou mesmo no uso do “broiler” para o preparo de alimentos – uma espécie de resistência na parte superior dos fornos, em sua maioria elétricos, que atinge temperaturas altíssimas (no Brasil, apenas alguns modelos de fogão mais caros e modernos possuem esse apetrecho atualmente);

C. Diferenças na construção de produtos condicionadas por diferenças climáticas;

→ Como não estou tratando de produtos, mas de preparações, poderia incluir nesta categoria os ingredientes, equipamentos e preparações típicos de cada

região que, mesmo com a globalização, podem ser difíceis de encontrar na cultura da língua de chegada. Uma receita britânica, por exemplo, que use *preserved green walnuts* (nozes verdes em conserva), ou uma americana que use o *buttermilk* (leitelho), ambos largamente disponíveis em seus países de origem, mas praticamente impossíveis de encontrar no Brasil atualmente, ou então uma receita brasileira que leve polpa de cupuaçu, ou use um cuscuzeiro em seu preparo.

D. Recomendações de produtos, específicas de cada cultura

→ Embora todos os manuais de estilo voltados para a redação de livros culinários sugiram que não sejam mencionadas marcas, alguns autores fazem questão de guiar seu leitor na escolha dos melhores produtos, especialmente nos livros voltados para usuários inexperientes. A substituição por nomes de outras marcas na língua de chegada deve ser discutida com o editor e, sempre que for possível e conveniente, a marca deve ser omitida, pois não há garantias de que a substituição surtirá os mesmos efeitos pretendidos pela menção da marca original no texto de partida. A tradução de um trecho como “*Tony Chachere’s creole seasoning*. This adds a sassy surprise to your favorite recipes; find its green can in the spice aisle.” pode ser bem complicada. Primeiramente porque o produto não existe no Brasil; em segundo lugar porque nem mesmo em seu rótulo é possível saber quais são os temperos usados em sua composição (“salt, garlic, red pepper and other spices”) – fica difícil, portanto, indicar uma substituição.

E. Textos padronizados, próprios de cada cultura;

→ O inglês possui textos mais ou menos padronizados para alertar o consumidor de um produto, ou o usuário que vai preparar uma receita, de que contém ingredientes potencialmente alergênicos (como nozes, glúten, leite), ou potenciais causadores de doenças (gemas ou carne cruas). Por exemplo: “The Department of Health advises that eggs should not be consumed raw. This book contains some dishes made with raw or lightly cooked eggs”. Esse tipo de texto não é comum em livros de receitas brasileiros – sua inclusão ou omissão deve ser discutida com o editor.

F. Orações declarativas próprias de cada cultura

→ Em inglês, por exemplo, é comum a estrutura: APPLY HEAT until APPEARANCE / CONSISTENCY, about TIME, p. ex.: “Cover and simmer until

almost tender, about 10 minutes.”, “bake until beginning to brown around the edges, about 20 minutes.”, “Add onion and bell pepper and sauté until golden and almost tender, about 4 minutes.”; em português, essa mesma estrutura declarativa é mais variável – a informação sobre o tempo pode ser omitida ou mesmo preceder a informação sobre a aparência / consistência desejadas: “Refogue em fogo baixo até o bacon começar a dourar.”, “Cozinhe até que o brigadeiro comece a desgrudar da panela.”, “Leve os pães ao forno por 45 minutos, ou até ficarem dourados.”;

G. Exemplos específicos de determinada cultura

→ Cada cultura escolhe suas analogias e metáforas de acordo com suas experiências do real. Se as receitas em português usam a “farofa” como analogia de consistência para o estágio inicial de uma massa não levedada (“Incorpore a manteiga gelada à farinha com a ponta dos dedos, até obter uma farofa fina”), em inglês usa-se “breadcrumbs” (pão esmigalhado / farinha de rosca grossa): “Rub the butter into the flour mixture until it resembles fine breadcrumbs”;

H. Ilustrações específicas de cada cultura

→ É possível que a foto de uma receita contenha ingredientes ou utensílios inexistentes ou cuja aparência seja distinta na cultura de chegada. O maracujá mais comum nos Estados Unidos, por exemplo, tem a casca vermelho-arroxeadada e é bem menor (cerca de 5 cm) que a variedade mais comum no Brasil, de casca amarela e podendo chegar a 15 cm; o abridor de latas típico americano caracteriza-se por um mecanismo giratório que se encaixa na borda da lata, pressionando-a entre dois discos cortantes, ao passo que o abridor mais comum no Brasil usa um sistema de alavanca; o double-boiler (duas panelas que encaixam uma na outra para preparos que exigem o banho-maria) não é facilmente encontrado no Brasil;

I. Métodos de medição específicos de cada cultura

→ Este é um grande dificultador das traduções no par de línguas inglês-português, já que os países de língua inglesa são os únicos que não usam o sistema métrico internacional. Todas as medidas em receitas são afetadas por essa diferença cultural: temperatura, peso, capacidade, extensão, e também tamanho de embalagens, tabletes, latas, garrafas, etc. Por exemplo, ao traduzir “2 cans (15 oz.) black beans, drained” é preciso saber que o tamanho de lata de

feijão mais comum nos Estados Unidos é de 15 1/2 oz, que equivale a 440g, ou 2 xícaras – um equivalente possível para essa oração seria “900g (4 xícaras) de feijão preto cozido, sem o caldo”;

- J. Falta de padronização terminológica entre empresas do mesmo ramo, e mesmo no interior de uma mesma empresa.

→ Na Culinária, os meios de divulgação em geral: livros, revistas, jornais, *sites*, programas televisivos, embalagens de produto, etc. são os grandes responsáveis pela difusão e até mesmo criação da terminologia da área. Há muito material traduzido (especialmente na Internet), alguns muito mal traduzidos, e ainda não foi feito um estudo terminológico sistemático da área como um todo no Brasil – trabalhos como o de Baccin (2003) são um grande avanço nesse sentido. Por exemplo, vide a variação encontrada em nomes de ervas (capim-santo, capim-cidreira, erva-cidreira, capim-limão); ou as variações regionais para nomes de ingredientes (macaxeira, mandioca, aipim) – um estudo socioterminológico descritivo abrangente seria o primeiro passo a ser dado antes que se possa propor qualquer tipo de ação normativa no Brasil nos setores necessários (mas preservando, sempre, a diversidade).

Ao que tudo indica, pode-se identificar na Culinária problemas semelhantes aos observados por Schmitt na tradução de manuais técnicos (1989 apud Azenha Jr., 1999: 79), o que corrobora, mais uma vez, o status da Culinária como área técnica de tradução. Vejamos agora se é possível encontrar exemplos também para as estratégias de tradução propostas por Barbosa (1990: 64-77) para os casos de não equivalência e equivalência parcial, que a autora subdivide em divergências de “Sistema Lingüístico”, “Estilo” e “Realidade Extralingüística”:

I. Divergências de “Sistema Lingüístico”

- *Transposição*: mudança de categoria gramatical.
Ex.: *pimenta-do-reino moída na hora* → *freshly ground black pepper*; *shaved parmesan cheese* → *lascas de queijo parmesão*;
- *Modulação*: reprodução da mensagem sob um ponto de vista diverso, que reflete uma diferença no modo como as línguas interpretam a experiência do real.
Ex.: *lettuce head* → *pé de alface*; *garlic clove* → *dente de alho*; *peel the avocado* → *tire a polpa do abacate da casca*;

- *Equivalência*: substituição de um segmento do original por um segmento funcionalmente equivalente.

Ex.: *seafood* → *peixes e frutos do mar*; *melt chocolate* *in a double-boiler* → *derreta o chocolate* *em banho-maria*

II. Divergências de “Estilo”

- *Omissão vs. Explicitação*: omitir elementos do original que, do ponto de vista da língua de tradução, são desnecessários ou repetitivos, ou ter que explicitá-los na direção inversa.

Ex.: *Mix sugar, flour, cocoa, butter and hot water* → *Misture* *o* *açúcar*, *a* *farinha*, *o* *chocolate em pó* e *a* *água quente*; *1 lb carrots, cut into 1/4 inch* *thick* *slices* → *1/2 kg* *de* *cenouras cortadas em rodelas de 0,5 cm*; *Wash and* *trim* *asparagus.* → *Lave os aspargos e* *descarte a parte mais dura dos talos*;

- *Compensação*: emprego de um recurso estilístico de efeito equivalente num ponto diferente do texto, quando não for possível reproduzi-lo no mesmo ponto.
→ Esta estratégia, em seu sentido estrito, pode ser usada em outros tipos textuais presentes em livros culinários, mas é pouco empregada na tradução de receitas (talvez ocorra apenas nos comentários apreciativos). Mas, se pensarmos em compensação de forma mais abrangente (i. e., ignorando a questão do “efeito estilístico”), poderíamos considerar, por exemplo, o apagamento de um aposto que explica um ingrediente na língua de partida e é desnecessário na língua de chegada e, em outro momento, o acréscimo de aposto para ingrediente na situação inversa, seja na mesma receita, seja em outra receita do livro, como um exemplo dessa estratégia;
- *Reconstrução de períodos*: re-dividir ou agrupar os períodos e orações do original.
→ Esta estratégia é muito usada na tradução de receitas, já que as receitas em português costumam ter períodos mais longos que as receitas escritas em inglês;
- *Melhorias*: omissão dos erros do original na tradução.
→ Esta estratégia deve ser usada na tradução de receitas sempre que possível, especialmente quando o erro do original compromete o preparo do prato. No exemplo a seguir, o problema estava no título (“pangriatata” não existe no

italiano – a autora errou na grafia de “pangrattato” que, nesta receita, refere-se a uma espécie de farofa feita de pão torrado moído com tomate seco):

Ex.: *Spaghetti with tomato pangriatata* → *Espaguete com pangrattato de tomate seco*

III. Divergências de “Realidade Extralingüística”

- *Transferência*: introdução de material textual da língua do original na língua de tradução. Tipos:

→ *estrangeirismo*: transferir para a tradução, marcando com aspas, itálico ou sublinhado itálico, vocábulos ou expressões do original que se refiram a conceito, técnica ou objeto desconhecido pelos falantes da língua de tradução (“empréstimo” para Vinay e Darbelnet)

Ex.: *1/2 cup cream cheese* → *1/2 xícara de cream cheese*;

→ *transliteração*: substituir uma convenção gráfica por outra.

→ **Não se aplica às receitas culinárias**

→ *aclimatação*: adaptação fonológica e morfológica dos empréstimos à língua que os recebeu (também chamado de “decalque” e “calque” por outros autores)

Ex.: *sandwich* → *sanduíche*; *cheeseburger* → *x burger*;

→ *transferência com explicação*: introdução de material textual da língua do original na língua de tradução acompanhada de uma explicação. Tipos:

- *notas de tradução*: notas de rodapé, notas de fim de capítulo ou notas / glossário ao final da obra;

Ex: *3 oz crème fraîche* → *100 g crème fraîche** (N.T. : *Especialidade francesa, é um tipo de creme de leite fermentado. Procure em casas de produtos importados, ou substitua por uma mistura de 4 partes de creme de leite fresco para cada parte de iogurte natural*)

- *explicação diluída no texto*: explicação entre vírgulas, entre travessões, entre aspas, entre parênteses; “equivalente cultural”

Ex: *1/2 cup maple syrup* → *1/2 xíc. (chá) de maple syrup (xarope de bordo)*; *2 Tbsp nam pla* → *2 colh. (sopa) de nam pla (molho oriental à base de peixe fermentado)*;

- *Explicação*: inserção da explicação que acompanharia o estrangeirismo, mas sem transferir o vocábulo do original.

Ex: *Sunday barbecue* → *Churrasco americano de domingo*; *Vegetarian BLT* → *Sanduíche de Tomate, alface e bacon vegetariano*;

- *Decalque*: tradução literal dos sintagmas ou tipos frasais da língua do original no texto da língua traduzida.

→ Esse fenômeno mereceria um estudo à parte. A influência do estilo e da fraseologia típica do inglês americano nas receitas brasileiras, em especial as traduzidas disponíveis em *sites* da Internet, é alarmante. Muitas vezes, o decalque vai caracterizar o “tradutês”¹¹⁸ – p. ex.: o uso de “chocolate semi-doce” (de *semi-sweet*) em vez de “chocolate meio-amargo”, “drene o excesso de líquido” (de *drain*), em vez de “escorra...”, ou “cebola vermelha” (de “red onion”) em vez de “cebola roxa”¹¹⁹. Eventualmente, com a repetição do uso, o “tradutês” pode tornar-se o termo / expressão consagrado – p. ex.: “creme (de leite) azedo” para “sour cream”. Um outro tipo de diferença extralingüística em Culinária que poderia, talvez, ser considerada como uma espécie de decalque pode ser visto no exemplo: “Deixar assar, passando de vez em quando o tempero reservado, cerca de 2 horas ou até que, enfiando um termômetro no centro da carne, no lugar mais grosso, ele registre 170°C.” – como já comentou-se, no Brasil não se usa termômetro para determinar o ponto de cozimento de carnes e, talvez, a estratégia de tradução mais adequada nesse caso fosse a adaptação;

- *Adaptação*: recriação de uma situação da língua do original que não existe na língua de tradução.

Ex.: **Canned beans**. Try Goya or Eden Organic. The azuki beans and the soybeans, as well as all the usual black and white ones, are tasty and quick. Just rinse and drain in a colander and proceed as if they were just cooked. → **Grãos e feijões cozidos** – tente os produtos cozidos no vapor da marca Vapza – há várias opções disponíveis. Há também os enlatados da Bonduelle, como o grão-de-bico, o feijão branco, a lentilha, etc. Enxágüe-os bem, escorra e use como se tivessem sido cozidos em casa;

¹¹⁸ *translationese*, em inglês; termo usado pela primeira vez por Gellerstam (1986) para descrever as diferenças de vocabulário observadas entre textos originalmente escritos em sueco e textos nesta língua que haviam sido traduzidos do inglês.

¹¹⁹ em Teixeira (2004: 149-180) citamos inúmeros exemplos problemáticos de decalque identificados na tradução do livro *A dieta de South Beach* (Agatston, 2003).

Como é possível observar, as estratégias de tradução para os casos de não-equivalência e equivalência parcial entre as línguas propostas por Barbosa, em sua grande maioria, podem ser aplicadas também à tradução de receitas. Essas estratégias visam a re-ambientação do texto traduzido ao seu novo espaço de recepção, buscando “equivalências” nos vários níveis lingüísticos e extralingüísticos, como bem observa Azenha Jr. (1999: 126): “(...) falar de ‘equivalência’ em tradução, inclusive em tradução técnica, implica falar também de acomodação a um outro espaço lingüístico-cultural, a um outro universo de valores”.

Os modelos apresentados aqui de categorização das divergências entre língua de partida e língua de chegada e respectivas estratégias de tradução nortearão o processo de estabelecimento de equivalências entre as Unidades de Tradução Especializadas levantadas nesta pesquisa, bem como a elaboração dos verbetes do dicionário voltado para o tradutor. Como veremos no Capítulo V, um dos diferenciais do verbete do dicionário eletrônico voltado para o tradutor que proporei é o campo de observações dedicado exclusivamente às questões relacionadas à tradução. Por ora, darei continuidade com a identificação de lacunas nos materiais de apoio disponíveis para o tradutor da área no Brasil.

2.4.2 Ferramentas de apoio ao tradutor da área

2.4.2.1 Dicionários

Até 2006, o tradutor da área de culinária no Brasil contava apenas com um dicionário bilíngüe no par de línguas português (brasileiro) – inglês. Atualmente são três, um deles ainda no prelo (com lançamento previsto para novembro de 2008):

- *Dicionário de Culinária e Termos Afins* (Moura, 2000)
- *Glossário de Gastronomia Português–Inglês / Inglês–Português* (Klie, 2006)
- *Vocabulário para Culinária* (Teixeira e Tagnin, 2008)

Em trabalho anterior (Teixeira, 2004: 19), discuti as características do dicionário de Moura (2000) e seus principais problemas como fonte de consulta para a tradução de receitas, tais como: equivalentes que não são de uso corrente em receitas brasileiras (p. ex.: *lump sugar* → *açúcar em tablete*; *sprinkle* → *aspergir*), equivalentes incorretos (*abacate* → *avocado*, *guacamole*), ou o uso de definições inteiras como

entrada na direção português-inglês (*amassar com a mão e abrir a massa com o rolo* → *knead*).

Publicado seis anos depois, o glossário de Klie (2006) tem uma proposta claramente didática, voltada para o ensino do inglês como segunda língua, que se deixa entrever tanto nas palavras da “Introdução” e da seção “Sobre a autora” quanto na organização interna da obra. Além do glossário nas duas direções, o livro conta com uma compilação de: i) expressões idiomáticas, ii) homófonos, iii) falsos cognatos, iv) estrangeirismos incorporados ao inglês, v) provérbios e ditos populares, vi) cortes de carne no Brasil e nos EUA, vii) metáforas e viii) tabelas de conversão de medidas e temperaturas.

Seções semelhantes podem ser encontradas em seu livro anterior, *Not just Hamburgers!* (Klie, 2003), uma edição bilíngüe de receitas americanas que traz, ao final, na seção “*More information about cooking / Mais informações sobre culinária*”, uma lista de i) *odd ingredients / ingredientes diferentes*, ii) *conversions / conversões*, *temperatures / temperaturas* e, logo em seguida, iii) “*Glossary and grammar in the kitchen / Glossário e gramática na cozinha*”, com: a) cognatos e falsos cognatos, b) expressões idiomáticas e, por fim, c) um glossário inglês–português, com cerca de 1.380 entradas. Ao que tudo indica, essas entradas com alguns acréscimos e a criação da direção português–inglês, geraram o glossário publicado em 2006 por Klie.

Ainda que o *Glossário de Gastronomia* de Klie forneça equivalentes para muitos novos termos, se for comparado com o outro único dicionário de Culinária bilíngüe disponível anteriormente (Moura, 2000), do ponto de vista da tradução culinária, há vários problemas na obra, especialmente na direção português–inglês.

Um desses problemas é a inconsistência no uso das convenções. Apesar de o livro apresentar uma lista de “Abreviaturas”, elas praticamente não são usadas. **GB** (Grã-Bretanha) e **US** (Estados Unidos), que indicam as variantes do inglês, parecem ser as mais usadas, embora com uma frequência mínima – a primeira ocorrência é na pág. 22, ainda que **abobrinha** > *zucchini* (= “courgette” em inglês britânico) e **abridor de lata** > *can opener* (“can” = “tin” em inglês britânico) ocorram logo na primeira página do glossário. As demais abreviaturas, como por exemplo, *ADJ* (adjetivo), *P.Ex.* (por exemplo), *SM-SF* (“substantivo ou locução com desinências diferentes para o masculino e o feminino”) praticamente não aparecem na obra (p. ex.: há uma ocorrência de *ADJ* na página 28 que, a propósito, causa bastante estranheza, já que a

categoria parece aplicar-se mais à palavra inglesa que ao equivalente oferecido em português: **com fome** *ADJ* > *hungry*).

A indicação de sinonímia, por sua vez, é feita de várias formas: com o uso de vírgula (como em **abóbora-moranga, moranga** > *pumpkin*), ponto-e-vírgula (como em **Bombril; esponja de aço; palha de aço** > *steel wool*) e parênteses (como em **Ajinomoto (glutamato monossódico)** > *MSG; monosodium glutamate* [sic]; *accent*). Aliás, os parênteses são usados para indicar várias relações diferentes na obra: ora servem para abrigar pequenas definições – como em **bacalhau (o peixe seco e salgado)** > *salt cod; salted cod fish*, ora para evitar a ordem inversa dos termos compostos, ou seja, a entrada segue a ordem alfabética do elemento que está após o parênteses (*alga* e *dendê*, nos exemplos a seguir) – (**uma das variedades de) alga marinha** > *dulse* e (**azeite-de-)dendê** > *red palm oil*, ou ainda para dar informações de ordem gramatical – como em **estripar** > *draw* (past *drew*, past participle *drawn*).

No entanto, o que mais me chamou a atenção foi a presença massiva de equivalentes cujo uso não seria adequado na tradução de receitas, por exemplo, ou que necessitariam de informações complementares (no caso de mais de um equivalente oferecido) para que fossem empregadas na produção do texto traduzido. Eis alguns exemplos: **bean pot** > caçarola refratária para feijão; (**navy**) **beans** > certa variedade miúda de feijão branco; **Barbados sugar** > certa variedade de açúcar bruto que é fino e úmido; **bagel** > determinado pãozinho em forma de anel; **batter** > massa crua de bolo; **biscuit cutter** > cortador de pão de minuto; **black walnut** > certa variedade de noz de sabor superior; **block ice** > gelo em barra, gelo em bloco; **bouillon** > sopa ou caldo ralo e transparente, normalmente de carne; **blueberry** > espécie de mirtilo; **wood alcohol** > espírito de madeira, álcool metílico; **mushroom** > champignon; cogumelo; funghi; (**to**) **stir-fry** > fritar mexendo sempre, à moda chinesa.

Uma terceira obra (Teixeira e Tagnin, no prelo) estará disponível no Brasil para os tradutores do par de línguas inglês-português até o final de 2008: o *Vocabulário para Culinária*, a ser lançado pela SBS como parte da série de dicionários técnicos *1001 Termos*, que conta com 19 títulos atualmente. Voltado explicitamente para a tarefa tradutória na direção inglês-português, o dicionário foi compilado de acordo com os princípios da Lingüística de Corpus, a partir do mesmo *corpus* que serve de base para a presente pesquisa (vide item 3.5 do Cap. III).

As entradas do dicionário foram levantadas “manualmente” (isto é, não usamos nenhum mecanismo de extração automática) no *corpus* de receitas originalmente

escritas em inglês com 1,5 milhão de palavras, usando o programa WordSmith Tools (Scott, 1997), a partir da observação das linhas de concordância geradas para cada uma das 300 primeiras palavras-chave do corpus¹²⁰. Em vista das limitações impostas pelo meio de divulgação da obra, foram privilegiadas as combinações de palavras mais recorrentes no corpus, em especial as formadas a partir de substantivos, verbos, adjetivos e advérbios. Os equivalentes *prima facie*, identificados por meio da consulta a obras de referência, à minha memória de tradução culinária e/ou à Internet, foram pesquisados no *corpus* de receitas originalmente escritas em português brasileiro, para determinar sua probabilidade de ocorrência em textos autênticos da área na língua de chegada.

Quando identificados possíveis problemas na tradução de uma entrada ou subentrada, informações adicionais foram acrescentadas ao verbete. Por exemplo: o verbete para “red onion” alerta o tradutor de que, em português, o equivalente não é “cebola vermelha” (uma tradução literal), mas “cebola roxa”; o verbete de “buttermilk” explica que esse produto não é facilmente encontrado no Brasil e sugere possíveis substitutos; o verbete de “red kidney beans” informa que um equivalente “possível” seria *feijão kidney*, e acrescenta: “o *feijão kidney* tem grãos grandes de coloração marrom-avermelhada bem escura; possíveis substitutos: *feijão jalo*, que também possui grãos graúdos, mas é de coloração mais clara, e *feijão vermelho* – escuro, mas de grãos pequenos”. Essas notas são identificadas por um símbolo especial dentro da obra, para chamar a atenção do usuário.

Nos casos em que as diferenças culturais mostraram-se muito grandes – por exemplo, ingredientes que existem numa cultura, mas não na outra; sistemas métricos; cortes de carne –, foram incluídas informações adicionais, como ilustrações, quadros de conversão de medidas, lista de temperos com respectivos nomes científicos e fotos, para que o usuário pudesse escolher a melhor forma de traduzir a informação adequadamente, em diferentes contextos. Para ampliar o conhecimento do consulente sobre a área, são feitas remissões para sinônimos, termos relacionados, variantes dialetais, homônimos e, quando pertinente, figuras, tabelas e imagens complementares. O sistema de remissivas também é destacado por um símbolo na obra, para facilitar a visualização.

¹²⁰ a terminologia da Lingüística de Corpus e seus principais métodos de pesquisa em corpora serão apresentados em detalhe no Capítulo III.

Um dos maiores diferenciais da obra é a abundância de exemplos de uso, com a respectiva tradução, em que o tradutor pode observar o contexto em que um termo ou fraseologia é empregado. Além disso, foram dadas informações, quando pertinente, sobre os colocados usuais. Por exemplo, o verbete para “bake” informa que os alimentos com que o verbo mais ocorre são: *loaves, bread, biscuits, cookies, cake, buns, rolls, potatoes, pies, pastries*; o verbete de “fold in” esclarece que, em geral, o termo co-ocorre com *egg whites*.

Outra inovação proposta é a inclusão de campos semânticos nas entradas do dicionário. Por exemplo: ao analisarmos as linhas de concordância para “slice”, “zest”, “fresh(ly squeezed)”, etc. observamos que ocorriam com várias frutas cítricas (p. ex.: *lemon, lime, orange, grapefruit*). Criamos, então, um verbete “semântico” para *CITRUS FRUIT* (em maiúsculas, para diferenciá-lo dos demais verbetes) em que registramos as fraseologias que mais co-correm com essas frutas no corpus, como mostra o excerto do verbete a seguir:

CITRUS FRUIT (n.)

☞ as frutas cítricas mais comuns em receitas em inglês são: *lemon* = limão siciliano; *lime* = limão (tahiti); *orange* = laranja; *grapefruit* = grapefruit / tor(a/o)nja. Vide verbete de cada fruta para mais especificidades

• **(finely) grated zest/rind of [NUMBER] [CITRUS FRUIT]** raspas (finas) de [NÚMERO] [FRUTA CÍTRICA]

E finely grated zest of 1 large orange = raspas finas de uma laranja grande

• **[CITRUS FRUIT] juice and zest** (o) suco e as raspas de [FRUTA CÍTRICA]

E juice and finely grated zest of 1 lemon = o suco e as raspas finas de 1 limão

• **[CITRUS FRUIT] pith** a parte branca do/a [FRUTA CÍTRICA]

E peel the oranges, removing as much of the white pith as possible = descasque as laranjas retirando o máximo possível da parte branca

• **[CITRUS FRUIT] rind** casca de [FRUTA CÍTRICA]

E Mix apples with raisins, lemon rind, sugar, cinnamon, and almonds = Misture a maçã com a uva-passa, a casca de limão, o açúcar, a canela e as amêndoas

• **[CITRUS FRUIT] segment(s)** gomo(s) de [FRUTA CÍTRICA] (sem a pele)

• **[CITRUS FRUIT] slices / sliced** rodela(s) [FRUTA CÍTRICA] / [FRUTA CÍTRICA] em rodela(s)

• **[CITRUS FRUIT] wedge(s)** [FRUTA CÍTRICA] (cortada) em cunhas / gomos

• **[CITRUS FRUIT], peeled and segmented** [FRUTA CÍTRICA], descascada e separada em gomos

• **[NUMBER] [CITRUS FRUIT], (grated) zest/rind only** [NÚMERO] [FRUTA CÍTRICA], somente as raspas / raspas de [NÚMERO] [FRUTA(s) CÍTRICA(s)]

E 3 oranges, grated zest only = raspas finas de 3 laranjas (...)

Figura 13: Exemplo de verbete “semântico” em Teixeira e Tagnin (2008)

Em seguida, criamos verbetes individuais para as frutas cítricas mencionadas no comentário explicativo, fazendo remissão para este verbete e, no caso das que

constavam entre as primeiras 300 palavras-chave, pesquisamos os padrões léxico-gramaticais específicos daquela palavra (p. ex.: *lemon*):

<p>lemon (n.) limão (siciliano)</p> <p>☞ variedade de limão mais difundida nos EU, tem a casca amarela e grossa – no Brasil, o limão mais comum é o tahiti, de casca verde e fina; embora seja mais fácil, atualmente, comprar o limão siciliano, em muitas receitas pode não ser necessário especificar.</p> <p>➔ para fraseologias, vide CITRUS FRUIT e squeeze of lemon/lime juice</p>	<p>lemon balm (n.) melissa</p> <p>lemon sole (n.) solha-limão / <i>linguado</i>* ☞ a <i>solha-limão</i> é um tipo de <i>linguado</i></p> <p>lemongrass (stalk / stick) (n.) talo de erva-cidreira / capim-santo / capim-limão ☞ no Brasil, usa-se a folha da planta para fazer chás, ao passo que na culinária asiática a parte usada é o talo (isto é, a parte branca logo acima da raiz)</p>
--	---

Figura 14: verbete “lemon” em Teixeira e Tagnin (2008)

O tradutor brasileiro conta também com duas obras traduzidas e adaptadas¹²¹ – *Dicionário prático de culinária* (Jones, 1996) e *Grande dicionário de culinária* (Dumas, 2006) –, um dicionário científico-químico de gastronomia traduzido do espanhol – *Léxico científico-gastronômico* (Fundação Alícia e ElBullitaller, 2007)¹²² – e três dicionários multilíngües que incluem as línguas português e inglês:

- *Tradutor gastronômico* (Engler, 2006)
- *Dicionário gastronômico* (Carli e Klotz, 2007)
- *Dicionário tradutor de gastronomia em seis línguas* (Saldanha, 2007)

O *Tradutor Gastronômico* de Engler (2006) foi publicado em parceria com a Abrasel (Associação Brasileira de Bares e Restaurantes). Segundo a autora me informou por e-mail, os direitos de publicação da obra foram vendidos ao Ministério do Turismo e, ao que tudo indica, o livro vem sendo distribuído apenas internamente – não é possível encontrá-lo para venda em nenhum dos *sites* de livrarias e sebos disponíveis na Internet¹²³. Por essa razão, Engler criou um *site*¹²⁴ com base no material

¹²¹ os originais não são propriamente dicionários – o nome da obra original de Jones é: *Cooking and Kitchen Skills*.

¹²² Elaborado a partir de pesquisas feitas pela Alícia, fundação de pesquisa científica catalã na área de gastronomia, e o laboratório culinário ligado ao restaurante elBulli, do catalão Ferran Adrià; contém cerca de 400 verbetes que focam especialmente os processos físico-químicos da gastronomia moderna.

¹²³ tentamos entrar em contato com a Abrasel por email várias vezes mas não obtivemos retorno.

¹²⁴ <http://myfoodtranslator.com/>

("Free instant food translator for beverage, cooking gadgets and tableware in English, French, Spanish, Portuguese, Italian and German.")¹²⁵.

Para ter uma idéia do conteúdo do *Tradutor Gastronômico*, já que não tive acesso à obra publicada, fiz um teste digitando "pimenta-do-reino moída na hora" na janela de busca e escolhendo o inglês como língua de chegada – nenhum equivalente é oferecido. Testei então a direção contrária digitando apenas "pepper" e escolhendo o português como língua de chegada. O equivalente "pimenta" aparece na janela de equivalente e um botão ("related terms") dá acesso aos termos relacionados, dispostos em forma de tabela:

black pepper	pimenta-do-reino
black pepper	pimenta preta
black pepper	pimenta do reino
cayenne pepper	pimenta-de-caiena
cayenne pepper	pimenta de caiena
green pepper	pimentão verde
green pepper	pimenta verde
hot pepper, red pepper, chilli	pimenta vermelha
pepper shaker	pimenteiro
red pepper, bell pepper	pimentão vermelho
sweet pepper	pimentão
Szechwan pepper, anise pepper	pimenta Cichuan, fagara
white pepper	pimenta branca
whole pepper	pimenta em grão
yellow pepper	pimentão amarelo

Tabela 3: termos relacionados a "pepper" no *Tradutor Gastronômico*

Como é possível observar pelo exemplo acima, o dicionário de Engler, assim como a maioria dos dicionários técnicos, não apresenta as informações de que o tradutor necessita para produzir um texto confiável na língua de chegada (p. ex.: diferenciação de homógrafos, indicação de variantes regionais, fraseologias, exemplos de contextos, ou indicação de frequência de uso nos casos de sinonímia). Embora a equivalência *pepper* → *pimentão* não apareça na janela do equivalente ("pimenta" é a única tradução oferecida para "pepper"), há várias ocorrências dessa correspondência dentre os "termos relacionados", sem qualquer observação quanto ao fato de serem homógrafos em inglês (vide, por exemplo, as várias ocorrências de *green* e *red pepper*) – numa receita, trocar *1/2 xícara (chá) de pimentão verde* por *1/2 xícara (chá) de*

¹²⁵ *Tradutor instantâneo gratuito para bebidas, equipamentos e utensílios de cozinha em inglês, francês, espanhol, português, italiano e alemão.*

pimenta verde, fará uma diferença considerável. Além disso, é possível ver no exemplo acima que não há consistência na sinalização das variantes (*red / green / yellow*) *bell* (U.S.) / *sweet* (U.K.) *pepper*, e nem indicações de grafia preferencial ou equivalente de uso mais corrente (o melhor equivalente para “black pepper” é “pimenta do reino” ou “pimenta-do-reino”? ou “pimenta preta”?).

O *Dicionário Gastronômico*, lançado em março de 2007, é de autoria do *chef* do Restaurante Cipriani do hotel Copacabana Palace, Francesco Carli, e de Eliane Klotz, diretora de marketing e eventos da Abaga (Associação Brasileira de Alta Gastronomia). Contém cerca de 1.700 termos em português e seus equivalentes em cinco idiomas: espanhol, inglês, alemão, francês e italiano. Assim como o dicionário de Engler, a obra consiste de uma lista. Direcionado a profissionais, estudantes e leigos no setor, surgiu de uma necessidade prática do dia-a-dia desses profissionais – nas palavras de Carli:

A idéia teve seu início quando encontramos dificuldades em elaborar menus e cardápios. A partir disso, começamos a pesquisa e surgiu esse novo desafio. A nossa preocupação é a de ajudar aos profissionais iniciantes, em traduzir para outros idiomas, os alimentos comuns entre nós e principalmente, os termos usados no dia-a-dia na cozinha¹²⁶

O dicionário de Saldanha (2007), escrito em colaboração com o chef Eugenio Mariotto, também parte do português e oferece equivalentes nas línguas: inglês, francês, alemão, italiano e espanhol. A direção língua estrangeira – português pode ser consultada via índice remissivo, que fornece a página e o número do verbete em português. A organização da obra é onomasiológica, ou seja, a autora optou por agrupar os 3.058 termos, numerados sequencialmente, em 32 campos semânticos, quais sejam:

• açúcares e afins;	• frutas, frutas secas e nozes;
• algas e plantas aquáticas;	• legumes, verduras e cogumelos;
• aperitivos;	• leites e iogurtes;
• aves e caças;	• manteigas;
• azeites,	• massas;
• óleos e gorduras;	• modos de preparo;
• bebidas;	• molhos;
• cafés e chás;	• outros ingredientes;
• caldos e sopas;	• ovos;
• cardápio;	• pães;

¹²⁶ <http://www.gastroonline.com.br/asp/noticia.asp?idtexto=4200> [acesso em 01/8/2008]

- carnes;
- cereais e grãos;
- comendo fora;
- condimentos e sementes;
- embutidos;
- ervas e especiarias;
- farinhas;
- peixes e crustáceos;
- saladas;
- sobremesas;
- termos culinários;
- utensílios de cozinha;
- vinagres.

É a obra mais completa e bem cuidada dentre as mencionadas, do ponto de vista da tradução. A propósito, no prefácio escrito por Caloca Fernandes¹²⁷, lêem-se comentários que, pela primeira vez, fazem referência explícita à tradução culinária:

Traduzi o meu primeiro livro de cozinha lá pelos idos de 1980. (...) Foi o meu primeiro contato com a dificuldade de uma tradução culinária que aceitei fazer sem supor o trabalho que teria pela frente. Nenhum dicionário conseguia orientar com precisão o termo português necessário para determinados procedimentos ao fogão – muito menos, os ingredientes. E olhe que eu tinha anos-luz de prática com receitas e, principalmente, com o rigor que eu exigia de mim próprio na redação delas. (...)

O dicionário mais do que bem-vindo de Roberta Saldanha não é o primeiro do mundo a nos prestar esse serviço, mas vai muito mais longe. Além da tradução pura e simples dos ingredientes, estende-se para expressões culinárias como *panar* (ou *empanar*), termos como *filé ao ponto para mal passado*, procedimentos como *servir bebidas* (não vá pensar que se trata de *to serve drinks...*), enfim, várias armadilhas que faziam das traduções um pesadelo tanto para o tradutor quanto para o leitor que delas dependia (...). (Saldanha, 2007: prefácio de Caloca Fernandes)

O tradutor pode recorrer também, em determinadas situações, às obras de referência monolíngües. Além do dicionário traduzido do francês *ABC da Cozinha*, de 428 páginas, editado por Victor Civita e publicado pela Editora Abril (1974)¹²⁸, fora de circulação, há quatro dicionários de culinária editados no Brasil:

- *Pequeno Dicionário de Gastronomia* (Gomensoro, 1999)
- *Pequeno Dicionário da Gula* (Algranti, 2000);
- *Glossário Técnico Gastronômico, Hoteleiro e Turístico* (Vieira e Cândido, 2000);

¹²⁷ tradutor, escritor e jornalista especializado em gastronomia.

¹²⁸ Adquiri este livro num sebo anos atrás, e até bem pouco tempo não tinha qualquer pista sobre sua procedência além do título e um selo do encadernador em que consta um telefone de São Paulo com apenas 6 dígitos (um indicativo de que se trata de obra não muito recente). O volume, reencadernado em capa dura, ao que parece, está incompleto – começa na página 5 e termina com a página 304, que contém verbetes da letra “L”. Em setembro de 2008 consegui localizar um exemplar completo, que data de 1974. Traduzida do francês *Tout à Vous - La Cuisine de A à Z*, a obra faz parte da coleção *Cozinha de A a Z*, em 7 volumes (2.134 páginas), também publicada pela Editora Abril na mesma época.

- *Dicionário-Almanaque de Comes & Bebes* (Fornari, 2001).

Em trabalhos anteriores, abordei os principais problemas dessas obras na qualidade de fontes de consulta para a tradução de receitas, destacando a falta de rigor metodológico, a cópia de outros dicionários (inclusive de língua geral e de edições portuguesas), a inclusão de termos pouco ou nada usados em textos correntes de culinária e até mesmo erros conceituais e julgamentos de valor (vide Teixeira, 2003 e 2004). Dicionários monolíngües só serão úteis ao tradutor na tarefa de decodificação – isto se, na direção inglês-português, for possível encontrar o equivalente para o termo ou unidade de tradução pesquisado nos dicionários bilíngües e multilíngües disponíveis. Vejamos, a seguir, em que outros tipos de fontes de consulta o tradutor pode se apoiar para traduzir receitas no par de línguas português-inglês.

2.4.2.2 Outros tipos de material impresso

Além de dicionários, o tradutor de culinária pode consultar outros materiais publicados no Brasil (e também importados, já que o acesso a essas obras é muito maior atualmente) para obter auxílio na tradução de receitas e textos culinários, como:

- livros de receitas escritos e publicados no Brasil, especialmente por editoras reconhecidas e especializadas na área, de preferência, com características semelhantes (público-alvo, tema, tipo de receita / ingrediente privilegiado, etc.) ao livro que se pretende traduzir;
- livros sobre ingredientes ou grupos de ingredientes específicos, como *Livro dos alimentos* (Gonsalves, 2001), *Ervas e especiarias na cozinha* (Junqueira, 1980), *Guia A-Z de Plantas: Condimentos* (Negraes, 2003);
- livros sobre técnicas de cozinha, como *Passaporte para o sabor* (Barreto, 2002), *400g: Técnicas de cozinha* (Kövesi et al., 2007), *Técnicas de cozinha profissional* (Sebess, 2008a), *Técnicas de confeitaria profissional* (Sebess, 2008b);
- obras bilíngües sobre a culinária americana ou brasileira (p. ex.: *Not just hamburgers!*, de Klie, 2003) e brasileira (o Senac, p. ex., possui vários títulos) e/ou sobre a culinária brasileira escritas em inglês, como, entre vários outros, *Eat smart in Brazil* (Peterson e Peterson, 1995);

- outras publicações especializadas e livros relacionados ao tema, como *Berinjela se escreve com j* (Melo, 1999), *Dicionário de termos de gastronomia francês / português* (Helene, 2006), *Guia de boas compras de gastronomia* (Mattar e Reis, 2007).

Dentre os diversos problemas de se consultar esse tipo de material durante a tradução podem ser citados:

- i. a dificuldade de se localizar contextos específicos – os índices remissivos são temáticos, e não lingüísticos, o que torna a pesquisa morosa e, não raro, infrutífera;
- ii. a confiabilidade dos materiais consultados – há uma série de traduções malfeitas e obras em português europeu sendo vendidas nas livrarias sem qualquer aviso ao leitor; além disso, em algumas obras bilíngües é evidente a presença de receitas que foram “escritas para serem traduzidas”, ou seja, usam ingredientes e processos para os quais é mais fácil encontrar equivalentes na língua de chegada¹²⁹;
- iii. “validade” do material consultado – com a globalização e o processo acelerado de internacionalização de ingredientes, os livros podem “perder a validade” para certos ingredientes / equipamentos rapidamente. Há 3 anos não era possível encontrar *tortillas* no Brasil com facilidade; atualmente, a marca Pullman® lançou o Rap10 (uma interessante aclimação de “wrap”, espécie de sanduíche feito com esse tipo de pão, cuja sonoridade imita “rapidez” em português).

2.4.2.3 Internet

Há uma infinidade de ferramentas de apoio ao tradutor da área na Internet. A mais usada delas (inclusive para localizar outras) é o motor de buscas Google¹³⁰, que pode ser instalado na barra de ferramentas do programa de acesso à Internet, facilitando as pesquisas¹³¹. Seu poderoso sistema de indexação permite as mais

¹²⁹ Por exemplo: Klie (2003: 37), na receita de *Cream of mushroom soup* inclui na lista de ingredientes “225 grams (1/2 pound) fresh mushrooms, finely chopped (preferably *Dinamarquês* or Shiitake mushrooms)”

¹³⁰ <http://www.google.com/intl/en/options/>

¹³¹ <http://toolbar.google.com/T5/intl/en/>

variadas buscas¹³², além de oferecer mecanismos específicos para buscas em livros, blogs, imagens, artigos acadêmicos, etc.

Ainda que o Google não tenha sido desenvolvido com vistas à pesquisa lingüística, é possível utilizá-lo para esse fim em vários momentos durante uma tradução. Por exemplo:

- para localizar equivalentes – em alguns casos, é possível encontrar equivalentes *prima facie* para termos na Internet. Por exemplo:
 - ao se digitar *mandioquinha english* na janela de buscas do Google obtém-se, entre os resultados retornados, equivalentes como: “white carrot”, “arracacha”, “Peruvian carrot” e “Peruvian parsnip”, além de variantes brasileiras como “mandiquinha-salsa”, “batata baroa”, “batata aipo” e “aipo do Peru”;
 - ao se digitar o nome científico ou parte dele na janela de busca e a língua de chegada (p. ex.: *rosmarinus english*) é possível encontrar equivalentes *prima facie* para plantas, frutas, legumes, verduras, peixes, etc.
- para comparar a frequência de uso de equivalentes de tradução *prima facie* quando há mais de uma possibilidade – p. ex.: ao testar o uso de dois equivalentes possíveis para “iogurte natural integral”, “plain whole milk yogurt” e “whole milk plain yogurt”, verei que o primeiro parece ser mais adequado, já que ocorre 15.000 vezes na Internet, em oposição a 603 ocorrências do segundo¹³³.
- para confirmar uma equivalência por meio da observação de imagens – quando o tradutor tem dúvida sobre a aparência de um determinado utensílio ou ingrediente pode solicitar uma busca somente em imagens no Google. Por exemplo: o equivalente *prima facie* encontrado para “feijão fradinho” em inglês é “black eyed beans”; o tradutor digita o termo na língua de partida e, em seguida, o termo na língua de chegada na janela de busca de imagens e confirma a semelhança de aparência entre os dois observando as fotos recuperadas.

O principal problema de se usar a Internet como ferramenta de apoio à tradução culinária é a confiabilidade das fontes de consulta, já que qualquer pessoa pode publicar textos em meio eletrônico (os culinários, especialmente as receitas, estão

¹³² vide <http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/Handouts.html> para dicas de pesquisa

¹³³ acesso em 02/8/2008

bastante disseminados em todo tipo de *site*). Por exemplo: o tradutor inexperiente que resolver testar “cebola vermelha” como equivalente *prima facie* de “red onion” (já que “red onion” não consta nem em Moura (2000), nem em Klie (2006) e nem em Saldanha (2007)), encontrará 12.500 ocorrências no Google¹³⁴. Se limitar a busca aos sites brasileiros – “cebola vermelha” site:.br – encontrará ainda 270 ocorrências (que pode considerar um número razoável de ocorrências). No entanto, o equivalente correto, “cebola roxa”, ocorre 20.300 vezes e, se limitada a busca apenas aos *sites* brasileiros, 9.180 vezes.

Além disso, a infinidade de possibilidades disponíveis na rede pode, na verdade, transformar-se num fator negativo, já que dificulta a localização de fontes confiáveis e tende a ocasionar dispersão e desconcentração do tradutor, que tem que examinar páginas e páginas de resultados em busca de contextos semelhantes àquele que precisa produzir na língua de chegada.

Para tentar minimizar esse problema, o Prof. Tony Berber Sardinha criou o KWIC Google¹³⁵ – uma ferramenta que organiza os resultados do Google centralizando a palavra de busca e faz uma lista das palavras que ocorrem com maior frequência à direita e à esquerda da palavra de busca num horizonte de até 5 palavras. Para usar o KWIC Google, é preciso: i) copiar a página de resultados do Google (para facilitar, o ideal é configurar o buscador para que retorne 100 ocorrências de cada vez); ii) colar no espaço reservado; iii) digitar a palavra que será usada para centralizar as ocorrências (= “nódulo”) no espaço reservado; iv) escolher o número de palavras para cada lado do nódulo que o programa deve considerar na contagem dos “colocados” (palavras que co-ocorrem à direita e à esquerda) e v) determinar a língua (inglês ou português), como pode ser visto na Fig. 15.

¹³⁴ acesso em 02/8/2008

¹³⁵ <http://www2.lael.pucsp.br/corpora/google/index.html>

Ajustes

Nóculo (digite apenas uma palavra):

Horizonte (esquerda e à direita, para cálculo dos colocados): 5 ▾

Língua em que foram apresentados os resultados pelo Google: Inglês ▾

Página de resultados do Google

Cole abaixo:

Search ResultsIn a Vegetarian Kitchen: Global FlatbreadsTurn the dough out onto a lightly floured board. Wash the bowl and lightly grease it. Thoroughly knead the dough, lightly sprinkling flour under it only as ...
vegkitchen.com/recipes/flatbreads.htm - 13k - Cached - Similar pages
Lots of Bread Recipes - Pg 7Turn the dough onto a lightly-floured surface. Knead until smooth and elastic. Transfer to a lightly-oiled bowl, cover with a damp towel and let rise in a ...
www.webvalue.net/recipes/bread7.htm - 17k - Cached - Similar pages
More Bread RecipesTurn onto a lightly floured surface and knead

sh, cgi, html: Tony Berber Sardinha.

Figura 15: configurações para usar o *KWIC Google*

Por exemplo: se o tradutor quer saber qual preposição usar na expressão “turn (the dough) ___ a lightly floured surface”, pode digitar na janela de busca do Google: *turn “a lightly floured surface”*. Após usar o KWIC Google (selecionando “lightly” para centralizar as ocorrências), poderá visualizar o resultado reordenado – a preposição “onto” é indicada como a de maior ocorrência à esquerda de “lightly” (vide Fig. 16).

Resultado semelhante, mas com opções mais refinadas de pesquisa, pode ser obtido com o WebCorp¹³⁶, um concordanciador *online* que usa a Internet como corpus. Entre as vantagens dessa ferramenta sobre o KWIC Google estão: i) oferece a opção de diferentes motores de busca (Google, AltaVista, MetaCrawler, AllTheWeb); ii) permite diferenciar maiúscula e minúscula; iii) permite escolher o formato dos resultados (html, txt, tabela); iv) oferece a possibilidade de exibir ou ocultar o endereço eletrônico do *site* que originou os contextos; v) permite escolher somente páginas que tenham sido atualizadas num determinado período de tempo; vi) permite limitar o número de resultados por *site* pesquisado, para que mais fontes sejam pesquisadas, entre outras. Infelizmente, o fato de ser extremamente lento torna-o inadequado para a tarefa tradutória, em que a rapidez é sempre um fator preponderante.

¹³⁶ <http://www.webcorp.org.uk/>

Figura 16: tela de resultado do KWIC Google

Outro *site* que permite uma busca mais refinada de contextos lingüísticos é o CorTec¹³⁷, Corpus Técnico do Projeto COMET (Corpus Multilíngüe para Ensino e Tradução), desenvolvido no Departamento de Letras Modernas da USP sob a coordenação da Profa. Dra. Stella E. O. Tagnin. O CorTec conta com vários *corpora* técnicos comparáveis (textos originalmente escritos em inglês e em português com características semelhantes) e ferramentas de pesquisa: concordanciador, listador de palavras e de agrupamentos de palavras (Tagnin, 2008). O Corpus de Culinária (Teixeira, 2005), constituído de receitas retiradas de *sites* da Internet, em sua primeira versão, contava com 368.227 palavras em inglês e 252.149 palavras em português. No segundo semestre de 2008, foi atualizado com o *corpus* resultante da presente pesquisa e contém agora cerca de 1,5 milhão de palavras em cada língua.

A pesquisa direta num *corpus* só de receitas facilita em muito o trabalho de pesquisa do tradutor, especialmente pela fiabilidade dos resultados. Mas, assim como no caso do KWIC Google e do WebCorp, os resultados exibidos não permitem a identificação rápida de padrões, especialmente por não permitirem a ordenação das linhas de concordância pelas palavras à esquerda ou à direita do nóculo. Por exemplo, se o tradutor estiver procurando um equivalente para “red onion” pode optar por

¹³⁷ http://www.fflch.usp.br/dlm/comet/consulta_cortec.html

examinar as linhas de concordância de “cebola*” no *corpus* de português do CorTec. Se examinar *en passant* as 1.075 ocorrências encontradas no *corpus* (primeira versão), será capaz de identificar algumas ocorrências de “cebola roxa” (vide Fig. 17).

Figura 17: tela de resultados do concordanciador do CorTec

No entanto, se um programa mais sofisticado de pesquisa de *corpus* fosse utilizado (como o WordSmith Tools), o tradutor identificaria com mais facilidade (vide Fig. 18) não só esse equivalente, mas os demais padrões de “cebola*”, usando ferramentas auxiliares, como o *Collocates* e o *Clusters* (vide item 3.3.4 do Capítulo III):

Concord - [CEBOLA: 6476 entries (sort: 1R,2R)]

N	Concordance	Set	Ta	Word	Nd
6233	de saquê mirin (licoroso) 1/2 xícara de óleo (aproximadamente) 1 cebola-roxa grande, em fatias bem finas 2 xícaras de camarão m				23.110
6234	Ervilha-torta e torta e cebola-roxa gratinadas Ingredientes: molho branco 4 colheres (s				3 1
6235	Ingredientes 1 radicchio pequeno 1 beterraba grande 1 cebola-roxa média cortada em rodela 150 g de uvas Rubi 10 ra				44 3
6236	ado alface crespa em pedaços pepino japonês fatiado sem pele 1 cebola-roxa média fatiada folhas de coentro folhas de hortelã 1 pi				99 4
6237	) de molho de tomate 500 g de lulas limpas e cortadas em anéis 1 cebola-roxa média ralada Modo de Preparo: Lave a cenoura e o				46 7
6238	vinagre e a cebola branca picada. Tempere. Coloque o agrião e a cebola-roxa numa tigela grande, tempere com um pouco do molho				1.228 c
6239	ção Colocar a manteiga e o azeite para derreter, depois colocar a cebola-roxa para fritar, colocando todos os cogumelos dentro por				111 0
6240	redientes: 500g de filés de linguado suco de limão quanto baste 1 cebola-roxa pequena 1 tomate pequeno sem sementes 1 pimenta				13 3
6241	ujás azedos pequenos 1/2 xícara de chá de salsinha picadinha 1 cebola-roxa pequena fatiada 2 colheres de sopa de mel 2 colher				23 2
6242	ido 1 pepino médio descascado e cortado em rodela de 1 cm 1 cebola-roxa pequena picada 1/2 xíc. (chá) de uvas passas pretas 1				149 7
6243	ido 1 pepino médio descascado e cortado em rodela de 1 cm 1 cebola-roxa pequena picada 1/2 xíc. (chá) de uvas passas pretas				149 1
6244	0 g de lentilhas verdes 1 cebola 1 talo de salsão, bem picado 1/2 cebola-roxa pequena, bem picada 1/4 col. (chá) de gengibre, des				765 a
6245	os de fraldinha de 200g sem gordura ou nervos 2 dentes de alho 1 cebola-roxa pequeno 16 azeitonas pretas 25 ml de azeite de oliva				19 0
6246	entões vermelhos médios 1 cenoura média Molho 1/4 xícara de cebola-roxa picada 1 colher (sopa) de azeite 1 colher (chá) de s				18.412
6247	fatias finas 3 pepinos grandes, sem casca Molho 1 xícara de cebola-roxa picada 1 xícara de água 1/2 xícara de suco de limão				13.705
6248	s de sopa de sal grosso 1/2 xícara de lentilhas 150 g de favas 1 cebola-roxa picada 3 cebolinhas picadas 1 colher de sopa de s				50 3
6249	pa) de gengibre ralado 2 dentes de alho picados 1 col. (sopa) de cebola-roxa picada 1 col. (chá) de talo de erva-cidreira 1 pimenta				70 c
6250	limão 1 colher (chá) de alcaparra escorrida 2 colheres (sopa) de cebola-roxa picada 1/2 colher (chá) de raiz-forte preparada (comp				288
6251	e a gosto COMPOTA DE CEBOLA ROXA 400 g de échalote ou cebola-roxa picada fina 200 ml de vinagre de vinho tinto 400 ml d				6.933 r
6252	picado 4 colheres (sopa) de salsinha picada 2 colheres (sopa) de cebola-roxa picadas 2 dentes de alho picados 1 pimenta dedo-de-				22 4
6253	engibre ralado 2 dentes de alho picadinhos 2 colheres (sopa) de cebola-roxa picadinha 1 colher (chá) de capim-limão em rodelinha				27.197
6254	NTAGEM Disponha as torradas nos pratos e passe a compota de cebola-roxa por cima. Coloque as fatias de foie gras, cortadas na				7.100 r
6255	ícaras de folhas de alface rasgada (roxa, crespa e lisa) 4 fatias de cebola-roxa separada em anéis sal e pimenta-do-reino a gosto.				1.036 c
6256	0051-Salada Arco-iris Ingredientes 1 cebola-roxa em rodela 2 cenouras raladas 1&				6 5
6257	* Sirva os chips com o creme quente, morno ou frio. Terrine de cebola-roxa, frutas secas, especiarias e molinho cremoso de				40.927
6258	, marroquinas e até sicilianas, mas sempre leva laranja, erva-doce, cebola-roxa, azeitona, tâmara, azeite e sal. * Como prato principal,				29.070
6259	mado, cortado em cubos 1 xícara de polpa de tomate 1 xícara de cebola-roxa, cortadas em rodela Sal e pimenta-do-reino, a gosto				27 4
6260	roquino com cordeiro (essa carne pode ser substituída por frango), cebola-roxa, fava, amêndoa e especiarias, que chega à mesa no t				29.094
6261	leite de coco, manga, banana, abacaxi, manjerição, coentro, alho, cebola-roxa, limão, tamarindo e capim-limão não são problema. S				20.985
6262	eixe esfriar. Transfira as lentilhas para uma tigela, junte o salsão, a cebola-roxa, o gengibre, o suco de limão, o molho de soja, o azeit				1.089 a
6263	s e texturas, como uma degustação. * O primeiro é uma terrine de cebola-roxa, que pode chegar à mesa quente ou fria, como sugeri,				39.622

Figura 18: linhas de concordância ordenadas pelas palavras à direita

O problema com essas ferramentas especializadas é que o tradutor tem que possuir o *corpus* de pesquisa – que, por razões de proteção dos direitos autorais, muitas vezes não está disponível para *download*, mas apenas para pesquisa. Além do mais, o exame de linhas de concordância provenientes de *corpora* comparáveis não resolve a questão da equivalência interlingual, que deve ser pesquisada nesse tipo de *corpus* usando ferramentas mais avançadas, como as que mostrarei no Capítulo III (a este respeito, vide também Tagnin, 2008).

2.4.2.4 Especialista

Especialistas em Culinária e Gastronomia são uma fonte de consulta importante para o tradutor, pois possuem o conhecimento teórico e prático da fenomenologia da área especializada. Como já discutido anteriormente (vide item 2.3.4), há vários tipos de níveis de especialização possíveis em Culinária e Gastronomia e, portanto, vários especialistas – cada qual vai dominar com mais confiança o jargão de sua esfera de atuação (culinária caseira, culinária profissional, alta gastronomia, culinária industrial, etc.).

No entanto, contar com a ajuda de especialistas no trabalho de tradução em geral é contra-producente, pois dificilmente haverá um profissional que esteja à disposição para dirimir dúvidas conforme elas forem surgindo no decorrer da tradução. Além disso, profissionais com diferentes formações e experiências práticas podem divergir de opinião quanto à terminologia da área. Portanto, tendo a chance de recorrer à opinião de especialistas, o tradutor deve tomar o cuidado de escolher o profissional de acordo com o público-alvo a que se destina seu texto de chegada, além de reconsiderar as sugestões dadas com cuidado do ponto de vista lingüístico, textual e cultural.

2.4.3 O que o tradutor de Culinária precisa

Pelo que foi exposto até aqui, existem várias lacunas nos materiais disponíveis para o tradutor da área que impedem ou, na melhor das hipóteses, dificultam o exercício da tarefa tradutória no par de línguas português brasileiro – inglês. Problemas semelhantes foram identificados em outras áreas especializadas¹³⁸ e em outros pares de língua por autores como Dancette & Réthoré (2000), Vintar (2000), Bowker e Pearson (2002), Avance de Souza (2002), Bevilacqua (2004), Gómez e Vargas (2004), Zanatta (2004), Gómez (2005), Salgado (2006), Carvalho (2007) e Fromm (2007). De modo geral, os estudos ressaltam os seguintes itens como necessários no dicionário técnico voltado para o tradutor:

- **Termos** – veiculam os conceitos da área especializada. As várias teorias terminológicas da atualidade têm dado conta, com maior ou menor êxito, da compilação de dicionários terminológicos monolíngües, mas pouco se têm feito em relação à terminologia bilíngüe, no que diz respeito às necessidades do tradutor. Faltam mais publicações de qualidade, especialmente nas áreas talvez "menos técnicas", como a Culinária. Infelizmente, o rigor metodológico pode implicar em morosidade na publicação (incompatível com as transformações aceleradas por que têm passado as áreas técnicas nos dias atuais);
- **Fraseologias especializadas** – descrevem o entorno imediato dos termos. Alguns autores têm estudado a fraseologia terminológica – p. ex.: Bevilacqua (1999), que estuda as Unidades Fraseológicas Eventivas (UFEs), centradas nos verbos, e Estopá (1999), que estuda as Unidades Terminológicas Poliléxicas

¹³⁸ vide item 3.4.3 do Capítulo III

(UTPs), cujo núcleo é um substantivo com valor de termo. No entanto, a presença das fraseologias especializadas em dicionários técnicos bilíngües ainda é tímida.

- **Contexto** – o item mais importante para o tradutor após a forma equivalente (Fromm, 2007: 62), pois é o que lhe permite produzir um texto natural na língua de chegada. Sua presença em dicionários terminológicos monolíngües é incipiente, e praticamente inexistente nas obras bi- e multilíngües;
- **Relações semânticas** – é importante para o tradutor saber como o conhecimento especializado de uma determinada área se estrutura, especialmente nos casos de equivalência parcial ou não-equivalência entre as línguas, pois só assim será capaz de produzir um texto explicativo ou nota de rodapé sobre o termo original. Essa informação pode ser dada em forma de definição formal, ou então sugerida por meio de esquemas prototípicos, árvores de domínio, etc., mas também pode ser encontrada em contextos definitórios.
- **Informações sobre o uso** – no caso de haver indicação de mais de um equivalente no dicionário voltado ao tradutor, é imprescindível também sinalizar as preferências de uso de cada um deles, seja em termos de frequência, seja com relação a uso restrito a determinadas regiões geográficas, níveis de especialização, etc. Isso não tem sido feito pelos dicionários técnicos monolíngües nem bilíngües – na área de Culinária, nenhum dicionário dá esse tipo de informação.
- **Informações extralingüísticas** – imagens podem não ser importantes em todas as áreas técnicas, mas o são na Culinária, constituindo-se, muitas vezes, no único recurso capaz de dirimir uma dúvida ou permitir uma descrição mais acurada para uma nota de tradução ou explicação apositiva.
- **Padronização textual** – ainda que a fraseologia especializada venha sendo estudada e conste de alguns dicionários técnicos produzidos recentemente pelos mesmos pesquisadores que se dedicam ao seu estudo teórico, não existe um tratamento adequado da padronização dos itens lexicais presentes nos textos especializados que não possuem valor de termo. Este é um dos objetivos deste trabalho: buscar compreender e descrever essa padronização.

Todas essas categorias estão previstas na Unidade de Tradução Especializada (UTE), justamente porque seu ponto de partida não é o termo, ou o conhecimento estruturado de uma área técnico-científica, mas os textos que veiculam esse conhecimento, e que são a matéria de trabalho primeira e última do tradutor. Ou seja, na prática tradutória, a distinção termo vs. não-termo parece não ser produtiva – o que, de certa forma, torna improdutivo também a distinção entre terminografia e lexicografia para esse público-alvo, como de fato tem sugerido alguns autores como Valente (2002), Pérez (2002), Temmerman (2003), Gómez (2005), Gómez e Vargas (2004) e Arribas Baño (2005).

A Lingüística de Corpus, a meu ver, é a abordagem teórico-metodológica que nos permitirá escapar dessas dicotomias para então focar o texto: ponto de partida e de chegada do tradutor. Dessa perspectiva, terminologia, cultura, fraseologia, padronização, recursos iconográficos, contextos, categorias gramaticais, etc. serão considerados na medida em que tiverem uma presença significativa em textos da área semelhantes àqueles que o tradutor terá que compreender e produzir na prática diária de sua profissão.

Disso decorre que um dos conceitos mais importantes de uma metodologia como essa é o de “corpus”, ponto central da Lingüística de Corpus, cujos principais fundamentos teóricos e práticos passo a explicar no capítulo que se segue.

This document was created with Win2PDF available at <http://www.daneprairie.com>.
The unregistered version of Win2PDF is for evaluation or non-commercial use only.

Capítulo III

Lingüística de Corpus e dicionários: mais uma metodologia, ou uma nova forma de observar textos especializados?

Neste capítulo, a Lingüística de Corpus é apresentada como abordagem teórica e metodológica, destacando as principais características que nos permitem apontá-la como adequada e necessária para a compilação de dicionários voltados para as necessidades de produção textual do tradutor técnico.

Primeiramente, faço um brevíssimo histórico da disciplina e apresento os conceitos principais da área que serão utilizados nesta pesquisa. Em seguida, relaciono a L.C. à teoria e à prática da tradução. Por fim, descrevo o processo de compilação dos corpora que servem de base para este trabalho – seu planejamento, a

coleta, limpeza e organização dos textos –, apresentando alguns dados quantitativos gerais.

3.1 Lingüística de Corpus: brevíssima introdução

A Lingüística de Corpus (L.C.), como seu próprio nome sugere, ocupa-se “da coleta e da exploração de corpora, ou conjunto de dados lingüísticos textuais coletados criteriosamente, com o propósito de servirem para a pesquisa de uma língua ou variedade lingüística” (Berber Sardinha, 2004: 3).

Refletir sobre o funcionamento e a natureza da língua por meio da observação de dados lingüísticos autênticos organizados em corpora não é, no entanto, uma invenção dessa abordagem teórico-metodológica – como afirma Rajagopalan (2007: 33), essa prática “é tão antiga quanto o surgimento do empirismo como método alternativo de fazer ciência”. O que muda com a Lingüística de Corpus, em termos concretos, é a forma como esses corpora são constituídos e analisados: por intermédio do computador.

A Lingüística de Corpus tem como um de seus marcos teóricos iniciais o texto de Quirk “Towards a description of English usage”, de 1960, em que o autor propõe seu projeto para o Survey of English Usage (SEU) – um corpus de 1 milhão de palavras, ainda em formato não eletrônico, que serviu de modelo para outros corpora que surgiram posteriormente (Berber Sardinha, 2004: 3; Rajagopalan, 2007: 32). A partir desse corpus, foram organizadas fichas, cada qual contendo uma palavra do corpus acompanhada de 17 linhas de contexto e sua respectiva categoria gramatical. Com base nos dados resultantes dessa pesquisa foi elaborada a extensa gramática *A Comprehensive grammar of the English language* (Quirk, Greenbaum, Leech e Svartvik, 1985), de 1.779 páginas. O corpus SEU foi informatizado somente décadas depois, em 1989 – a digitalização da parte falada ficou pronta antes, e é conhecida atualmente como *London-Lund Corpus* (Berber Sardinha, 2004:4).

O corpus *Brown*, primeiro corpus computadorizado, começou a ser compilado por W. Nelson Francis e Henry Kucera, no início da década de 1960. A exemplo do SEU, continha 1 milhão de palavras distribuídas em 200 textos jornalísticos com 5.000 palavras cada, em 15 categorias. Sua criação é concomitante ao aparecimento dos computadores *mainframe*, presentes apenas em algumas universidades, à época. As

dificuldades para informatizar textos, então, eram enormes – eles eram inseridos no computador por meio de cartões perfurados.

Na mesma época foi lançado o primeiro livro de Noam Chomsky, *Syntactic Structures* (1957). Segundo Rajagopalan (2007: 24), a partir da publicação desse livro e da crítica feita por Chomsky, em 1959, ao livro *Verbal Behavior* de Skinner, “a palavra empirismo tornou-se amaldiçoada (junto com a palavra behaviorismo), pois (...) o espírito do racionalismo cartesiano começou a varrer a Lingüística”. Rajagopalan (*id. ibid.*) ilustra bem esse “espírito” citando uma passagem de *Diálogos*, em que Platão comenta, a respeito de Sócrates, o “Pai da Filosofia Ocidental”, que ele não se contentava enquanto não obtivesse de seu interlocutor uma definição (de “piedade”, por exemplo):

(...) meros exemplos de pessoas piedosas que o homem simples ateniense enumerava em resposta à pergunta insistente do filósofo não eram suficientes para iluminar o conceito da piedade propriamente dita. Para Sócrates, os exemplos (em nossos tempos, os dados) eram apenas ilustrativos. O que importava mesmo eram os conceitos que somente podiam ser apreendidos por meio da intuição e não mediante consulta empírica. (Rajagopalan, 2007: 24).

À margem da teoria racionalista chomskyana (chamada, ao longo dos anos, de Gramática Transformacional-gerativa, Teoria Padrão, Teoria x-barras, Minimalismo ou simplesmente Gerativismo), a Lingüística de Corpus continuou evoluindo, especialmente na Europa, e em particular na Grã-Bretanha, pelos 30 anos seguintes, ainda que timidamente. Autores considerados hoje os maiores expoentes da área, como John Sinclair e Geoffrey Leech, continuaram suas pesquisas, publicando trabalhos importantes entre 1960 e o final da década de 1970. O cenário começou a mudar definitivamente a partir da década de 1980, com a difusão em larga escala dos computadores pessoais e o lançamento do projeto COBUILD¹³⁸ – parceria pioneira estabelecida entre a Universidade de Birmingham e a editora Collins, com vistas à produção do primeiro dicionário de língua (inglês) compilado segundo os preceitos da L.C., o *Cobuild English Dictionary* (Sinclair *et al.*, 1987).

Outras parcerias entre universidades e instituições privadas promoveram o avanço e a popularização da Lingüística de Corpus. Na Europa, as editoras de

¹³⁸ Collins Birmingham University International Language Database

dicionários de língua geral tornaram-se grandes aliadas dos projetos lexicográficos desenvolvidos por lingüistas de corpus (p. ex. Longman, Oxford), ao passo que, nos Estados Unidos, as empresas de telecomunicação e informática foram as que mais investiram no setor (p. ex.: Xerox, Microsoft, Canon), o que justifica a forte associação que a L.C. tem nesse país com o PLN e a Lingüística Computacional (Berber Sardinha, 2004: 6).

Os principais diferenciais dos estudos em L.C., que vêm atraindo a atenção de pesquisadores dos mais variados campos, são: i) as pesquisas são empíricas: analisam os padrões de uso em textos autênticos, produzidos naturalmente; ii) são utilizadas grandes quantidades de dados, criteriosamente organizados (= corpus); iii) o computador é usado extensivamente nas análises; iv) os estudos empregam técnicas quantitativas e também qualitativas (Biber, Conrad e Reppen, 1998: 4).

As diferentes pesquisas em Lingüística de Corpus que se desenvolveram nesses últimos 30 anos na área poderiam ser agrupadas, de acordo com Berber Sardinha (2004: 38), sob três paradigmas principais: um mais informal, baseado na análise de linhas de concordância – é o que concentra o maior número de pesquisas na área; e dois mais quantitativos, que fazem uso extensivo de técnicas estatísticas avançadas (como os modelos *log-linear* e de *Markov*). Quanto às áreas de pesquisa privilegiadas, há pesquisadores que se dedicam à compilação de corpora, à descrição da linguagem e ao desenvolvimento de novas ferramentas de exploração de corpora eletrônicos, mas o maior número de trabalhos concentra-se na vertente aplicada da L.C., por exemplo:

- Lexicografia e Terminologia: construção de dicionários, glossários, bases de dados terminológicos, etc.;
- Aprendizado e ensino de língua estrangeira: observação de exemplos autênticos da língua em uso, produção de material didático;
- Lingüística Computacional e Processamento da Linguagem Natural: desenvolvimento de corretores ortográficos e gramaticais, tradução automática, reconhecimento de voz, etc.;
- Tradução: observação de exemplos de uso natural em textos especializados produzidos naturalmente, contextos definitórios, equivalentes tradutórios, etc.

A relação da L.C. com a Tradução e a compilação de dicionários técnicos será examinada mais detalhadamente no decorrer deste capítulo. Mas, antes, vejamos

alguns dos princípios teóricos e metodológicos fundamentais da L.C. que serão usados neste trabalho, a começar com uma discussão da diferença entre usar a L.C. como metodologia e como abordagem para o estudo de textos especializados.

3.2 Lingüística de Corpus como abordagem e como metodologia

Um aspecto que venho frisando desde a Introdução deste trabalho é a diferença entre usar a L.C. como abordagem teórica, fundamentada por uma visão empirista e probabilística da linguagem, e usá-la como metodologia de análise. A discussão dessas diferentes perspectivas de estudo passa pela questão da incorporação do computador à pesquisa lingüística.

De modo geral, o computador passou a fazer parte da maioria das pesquisas lingüísticas de base empírica nas últimas décadas – muitas vezes, com o intuito de automatizar tarefas antes enfadonhas, cansativas e propensas às interferências de erros humanos, quando não impossíveis (como contar e/ou analisar manualmente todas as ocorrências de um dado item lexical num corpus de 1 milhão de palavras).

Para o pesquisador que usa a L.C. como abordagem, no entanto, a incorporação do computador à análise lingüística representou uma contribuição muito maior: permitiu que a visão da linguagem como **probabilidade**, que tem em Michael A. K. Halliday seu maior expoente, voltasse a ser explorada – agora com um número muito maior de evidências e contando com o apoio de ferramentas informatizadas robustas e precisas.

Para autores como Hoey (1993) e Stubbs (1996: 231), a importância da incorporação do computador à análise lingüística pode ser comparada à importância da invenção do microscópio para a biologia pois, não só ampliou, mas transformou o conhecimento que se tinha até então da língua, permitindo observar fatos que não podiam ser vistos “a olho nu”. Ou seja, não se trata apenas de ter uma quantidade maior de dados lingüísticos à disposição. As ferramentas computadorizadas de análise textual permitem **observá-los de uma outra perspectiva**, evidenciando **novos fenômenos**.

Tognini-Bonelli (2001: 2) assim explica a diferença de trabalhar com evidências empíricas identificadas por meio de corpora eletrônicos:

(...) a text exists in a unique communicative context as a single, unified language event mediated between two (sets of)

participants; the corpus, on the other hand, brings together many different texts and therefore cannot be identified with a unique and coherent communicative event; the citations in a *corpus* – expandable from the KWIC format to include *n* number of words – remain fragments of texts and the significant elements in a *corpus* are the patterns of repetition and patterns of co-selection. This difference entails a different “reading” of the two: the text is to be read horizontally, from left to right, paying attention to the boundaries between larger units such as clauses, sentences and paragraphs. A corpus, examined first in KWIC format with the node word aligned in the centre, is read vertically, scanning for the repeated patterns present in the co-text of the node.¹³⁹ (Tognini-Bonelli, 2001: 2-3)

Essa forma de abordar a linguagem é chamada, na literatura, de *corpus-driven approach* (abordagem direcionada pelo corpus). Surgiu como uma crítica às pesquisas “baseadas em corpus” (*corpus-based*), isto é, aquelas em que os corpora (e as ferramentas informatizadas de exploração) são utilizados apenas para confirmar teorias pré-concebidas, ou como repositórios de exemplos – “fish ponds” (pesqueiros), nas palavras de Sinclair (Hanks, 2008)¹⁴⁰.

Usar a L.C. como abordagem teórica significa seguir o lema de Sinclair: “trust the text” (confie no texto). Essa visão mais empirista da linguagem representa atualmente uma das correntes de pesquisa mais significativas dentro da L.C. (Berber Sardinha, 2004: 36). Para Sinclair, não se deve partir de nenhuma categorização apriorística da linguagem, mas observar o que de fato é usado pelos falantes, buscando evidências na frequência e na co-ocorrência de itens lexicais do corpus para relacioná-las, depois, a outras instâncias da comunicação.

Foi observando as regularidades da linguagem que Sinclair identificou padrões idiomáticos, isto é, expressões pré-fabricadas de linguagem que são selecionadas pelo falante como um bloco indecomponível – fenômeno que ficou conhecido como *idiom*

¹³⁹ (...) um texto existe num contexto comunicativo único como um evento lingüístico singular mediado por dois (grupos de) participantes; o corpus, por outro lado, reúne muitos textos diferentes e, portanto, não pode ser identificado como um evento comunicativo único e coerente; as citações em um corpus – que podem ser expandidas a partir das linhas de concordância para conter *x* palavras – continuam sendo fragmentos de textos, e os elementos significativos em um corpus são os padrões de repetição e os padrões de co-seleção. Essa diferença implica uma “leitura” diferente dos dois: o texto deve ser lido horizontalmente, da esquerda para a direita, prestando-se atenção às fronteiras de unidades maiores como sintagmas, sentenças e parágrafos. Um corpus, examinado primeiramente por meio das linhas de concordância com a palavra de busca centralizada, é lido verticalmente, buscando-se padrões repetidos no co-texto da palavra de busca.

¹⁴⁰ O termo *Corpus-based approach*, por outro lado, é a denominação preferida por autores, como Douglas Biber, para identificar a pesquisa que usa a L.C. como abordagem e não como mera metodologia (Berber Sardinha, 2004: 37-38).

principle (princípio idiomático). Esses padrões, para Sinclair, são uma evidência da impossibilidade de se separar léxico e gramática – um co-seleciona o outro; ou, em outras palavras, a escolha de um item lexical ou categoria gramatical interfere na probabilidade de outros itens lexicais e categorias gramaticais ocorrerem naquele mesmo contexto.

É nesse ponto que a L.C., como abordagem teórica da linguagem, aproxima-se da teoria Sistêmico-funcional hallidayana, que também defende a existência de uma léxico-gramática, composta de padrões cuja característica mais importante é a probabilidade que têm de ocorrerem em determinados contextos. Essa probabilidade depende da situação de comunicação em que esses contextos se inserem, que, por sua vez, é condicionada por fatores como: o tópico da interação, as relações entre os participantes, o meio de divulgação e o gênero.

Systemic linguists chart their analyses by diagramming the choices language users can make in a given setting to realize a particular linguistic product. The available choices depend on aspects of the context in which the language is being used. Choices can be charted on different levels, or strata, of language.

(...) Because language is rich and multi faceted, the relevant aspects to be highlighted in linguistic research must, in part, be determined by the research objectives. The "strata" perspective of systemic linguistics allows for this flexibility across research needs while maintaining its fundamental definition of language as a resource for meaning potential¹⁴¹. (Chapelle, 1998)

Ainda que Halliday não se autodenomine um lingüista de corpus (como o faz Sinclair), e que suas análises possam prescindir do uso de corpora informatizados e das ferramentas de exploração privilegiadas pela L.C., a Lingüística Sistêmico-funcional “serve como arcabouço teórico maior no qual ela [a L.C.] se pode incluir” (Berber Sardinha, 2004: 35). Este mesmo autor, em artigo recente (Berber Sardinha, 2008), ao destacar a noção de “registro” na teoria Sistêmico-Funcional, discute a

¹⁴¹ Os lingüistas sistêmico-funcionais delineiam suas análises por meio da diagramação das escolhas lingüísticas que os usuários têm à disposição numa determinada situação para realizar um produto lingüístico em particular. As escolhas disponíveis dependem de aspectos do contexto em que a linguagem está sendo usada. Essas escolhas podem ser delineadas em diferentes níveis (ou strata) da linguagem. (...) Como a linguagem é rica e multifacetada, os aspectos relevantes a serem destacados na pesquisa lingüística devem ser determinados, em parte, pelos objetivos do estudo. A perspectiva dos “strata” adotada pelos lingüistas sistêmico-funcionais permitem essa flexibilidade para acomodar as necessidades de pesquisa ao mesmo tempo em que mantém a definição fundamental de linguagem enquanto recurso significativo potencial.

relação entre léxico-gramática, gênero e probabilidade – o que, a meu ver, diz respeito diretamente ao uso da L.C. como abordagem adequada para estudar o texto técnico. Ele observa que os padrões de probabilidade variam de acordo com uma série de fatores – a probabilidade observada para um padrão lexico-gramatical num dado registro pode não ser a mesma em outros registros e/ou na língua geral (Berber Sardinha, 2008). O autor se apóia em Halliday (1993: 24), que define registro como “the resetting of probabilities in the lexis and in the grammar”¹⁴².

Isto é, em vez de se pensar em gênero e tipologia textuais como conjuntos mais ou menos definidos pelas peculiaridades do vocabulário encontrado nos textos que os representam, poderíamos, inversamente, pensar em gênero e tipologia textuais como **definidores do contexto de ocorrência e co-ocorrência dos padrões léxico-gramaticais** de uma língua, contexto este que determinará a probabilidade maior ou menor de esses padrões serem usados para expressar os sentidos que ocorrem tipicamente naquele âmbito da interação humana. Isto implica que o tradutor, somente após relacionar os padrões léxico-gramaticais do texto original às variáveis que condicionaram sua produção na língua de partida será capaz de avaliar quais seriam os padrões léxico-gramaticais na língua de chegada capazes de estabelecer, no contexto de produção textual da tradução, um valor de “equivalência”, seja esse valor alcançado por meio de recursos expressivos do nível léxico-gramatical, semântico, interpessoal, textual ou cultural. E para produzir uma “boa” tradução ele tem que conhecer, entre outras coisas, os padrões léxico-gramaticais mais prováveis para expressar aquela informação em textos típicos da língua de chegada.

a ‘good’ translation is a text which is a translation (i.e. is equivalent) in respect of those linguistic features which are most valued in the given translation context (...) and perhaps also in respect of the value which is assigned to the original (source language) text¹⁴³. (Halliday, 2001: 17)

Como bem observa Stubbs (2001: 14): “It is not the words which tell you the meaning of the phrase, but the phrase which tells you the meaning of the individual words within it”. Analogamente, poderíamos dizer que não são os termos que apartam

¹⁴² a reformulação das probabilidades no léxico e na gramática

¹⁴³ uma “boa” tradução é um texto que é uma tradução (i.e. é equivalente) no que respeita às características linguísticas que são mais valorizadas num dado contexto de tradução (...) e, talvez, também com relação ao valor que é dado ao original (texto na língua de partida).

a tradução de textos técnicos da tradução de outros tipos de textos, mas as condições de produção e recepção do texto técnico na língua de chegada, pois contribuirão para determinar quais padrões léxico-gramaticais tem maior probabilidade de ocorrer, ou ocorrem tipicamente, naquele tipo de contexto. Disso concluo, com Berber Sardinha, que “**o tradutor traduz gênero (ou outra forma social mais ou menos estável) e não termos**” (comunicação pessoal).

O uso da L.C. como **abordagem** no estudo dos textos técnicos, e não apenas como um conjunto de métodos computadorizados para a exploração de textos, é, portanto, o posicionamento mais acertado quando se tem em vista produzir um material de apoio voltado para o tradutor-produtor-de-textos de um gênero e tipo textuais bem específicos: a receita culinária. Isso implica que olharei para os textos técnicos do gênero e tipologia escolhidos não como repositórios de termos, mas como um sistema de probabilidades lingüísticas em que, dependendo do contexto de produção textual, certos padrões léxico-gramaticais são selecionados tipicamente, em detrimento de outros, também disponíveis no sistema da língua como um todo e com probabilidades diversas de ocorrência em outros gêneros e tipologias textuais.

Para abordar o texto dessa perspectiva é preciso seguir certos princípios, apresentados nos itens a seguir, que vêm permitindo considerar as pesquisas feitas em L.C., em consonância com o que afirmam Quirk *et al.* (1985), Leech (1992) e Hoey (1997), “não somente como uma nova metodologia emergente para o estudo da linguagem, mas uma nova empreitada de pesquisa e, na verdade, uma nova abordagem filosófica” (Leech, 1992: 106 *apud* Berber Sardinha, 2004: 37).

3.3 Princípios basilares da Lingüística de Corpus

3.3.1 Corpus

A palavra “corpus”, em seu sentido lato, pode ser usada para referir-se a qualquer conjunto de dados, documentos ou informações sobre determinada matéria. Em L.C., no entanto, “corpus” tem um significado restrito, que reflete os princípios que norteiam a compilação e a exploração deste que é o objeto de estudo privilegiado de dessa abordagem.

Berber Sardinha (2004: 18-19) elenca vários critérios a serem considerados na compilação de um corpus para que possa servir de fonte de pesquisas em L.C., que poderiam ser resumidos em:

- **Autenticidade:** os textos devem ter sido produzidos naturalmente, ou seja, sem o propósito explícito de serem usados para determinada pesquisa;
- **Naturalidade:** os textos devem ter sido produzidos por falantes nativos – se os corpora forem constituídos de textos produzidos por falantes não nativos (p. ex., corpora de textos produzidos por aprendizes de uma segunda língua, corpora de textos traduzidos), essa informação deve ser explícita;
- **Representatividade:** sem perder de vista o fato de que os corpora são, por natureza, finitos e não-exaustivos, deve-se fazer o possível para que sejam representativos da linguagem ou variedade lingüística para a qual servirão de fonte de evidências – o que, em geral, significa que são extensos (esse critério será discutido com mais vagar no item 3.3.2);
- **Formato:** os corpora devem estar em formato eletrônico, de forma que possam ser explorados por ferramentas computacionais diversas;
- **Propósito:** os corpora devem ter sido coletados com o propósito específico de servirem de fonte para uma determinada investigação lingüística, ou seja, a reunião dos textos que o compõem não é aleatória – deve seguir critérios bem definidos e previamente determinados.

Mona Baker, estudiosa da Tradução, propôs uma definição de corpus que contempla a maioria desses princípios:

a collection of texts held in machine-readable form and capable of being analysed automatically or semi-automatically in a variety of ways (...) it is put together for a particular purpose and according to explicit design criteria in order to ensure that it is representative of the given area or sample of language it aims to account for¹⁴⁴ (Baker, 1995: 225).

Para Berber Sardinha (2004: 18), a definição que menciona os pontos mais importantes que caracterizam um corpus é a de Sánchez e Cantos (1996):

¹⁴⁴ *uma coleção de textos disponíveis eletronicamente e organizados de modo a permitir análises automáticas e semi-automáticas de diversos tipos (...) é coletado com uma finalidade específica e de acordo com critérios bem definidos de construção, de modo a garantir que seja representativo de uma dada área ou variedade lingüística que se propõe descrever.*

um conjunto de dados lingüísticos (pertencentes ao uso oral ou escrito da língua, ou a ambos), sistematizados segundo determinados critérios, suficientemente extensos em amplitude e profundidade, de maneira que sejam representativos da totalidade do uso lingüístico ou de algum de seus âmbitos, dispostos de tal modo que possam ser processados por computador, com a finalidade de propiciar resultados vários e úteis para a descrição e análise. (Sánchez e Cantos, 1996).

Pode-se observar que ambas as definições não abordam a questão da naturalidade e da autenticidade dos textos, que é da maior importância para este trabalho, pois ele preocupa-se com a produção textual. É certo que uma definição de corpus que mencione a questão da naturalidade como critério de inclusão de textos em corpora (isto é, que os textos tenham sido produzidos por falantes nativos), acabaria por desconsiderar os corpora de aprendizes, de tradução, de textos produzidos por pessoas disléxicas, etc. Mas a autenticidade (ou seja, os textos não podem ter sido produzidos para fins da pesquisa que se pretende desenvolver), a meu ver, deveria ser mencionada.

Outra questão que parece não ser abordada por essas definições de corpus é a fragilidade do conceito de representatividade (vide item 3.3.2). Como bem coloca Berber Sardinha (2004: 19), “Cabe perguntar: representativo do quê e para quem?”. Como o corpus é “uma amostra de uma população cuja dimensão não se conhece” (*op. cit.*: 23), não é possível estabelecer um tamanho ideal para essa amostra, de forma que não se pode afirmar, a princípio, que seja representativa de qualquer população. A representatividade de um corpus, portanto, no estado atual das pesquisas em L.C., consiste mais num “ato de fé” (Leech, 1991: 27 *apud* Berber Sardinha, 2004: 25).

Considera-se que o corpus deve ser grande o bastante (quanto maior, melhor) para garantir uma alta probabilidade de ocorrência dos fenômenos que se pretende estudar, de modo que possam ser feitas generalizações (por exemplo: um corpus para o estudo dos pronomes pessoais em textos técnicos certamente não precisaria ser tão extenso quanto um corpus destinado ao estudo lexicográfico exaustivo de uma dada língua). Ainda que a escassez de recursos disponíveis sobre um assunto, ou a dificuldade de acesso a textos de uma determinada matéria possa justificar, em alguns casos, a construção de um corpus “oportunistico” (isto é, que contém aquilo que foi possível coletar, diante dessas dificuldades), certamente não autoriza a compilação de

um corpus “tendencioso” (*biased*), isto é, que contém apenas o que se espera comprovar ou refutar naquele estudo em particular.

Em vista disso, proponho trabalhar com um conceito de corpus que considere as questões acima e que não esteja tão centrado nos critérios usados para a compilação de corpora de língua geral, destinados à descrição da linguagem como um todo. A definição que melhor descreve esses aspectos na literatura, ainda que seja curta e simples, é a de Bowker e Pearson (2002: 9), que também trabalham com textos especializados: “A corpus can be described as a large collection of **authentic** texts that have been gathered in electronic form according to a specific set of criteria”¹⁴⁵ (grifo meu). Talvez, essa definição pudesse ser reescrita da seguinte forma para torná-la mais adequada à tarefa a que me proponho neste trabalho:

corpus é um conjunto de **textos autênticos** disponível em formato eletrônico, coletado com uma finalidade específica, segundo critérios previamente estabelecidos, e **extenso o bastante** para permitir que sejam feitas **generalizações sobre o(s) aspecto(s)** dessas línguas ou variedades lingüísticas que se pretende estudar, isto é, que possa ser considerado uma representação dessa língua ou variedade lingüística.

3.3.1.1 Tipos de corpus

Os corpora e os textos que os compõem podem ser classificados de acordo com vários critérios, como pode ser observado nas propostas disponíveis na literatura da área (vide, por exemplo: Atkins, Clear e Ostler, 1992; Biber, Conrad e Reppen, 1998: parte IV; Berber Sardinha, 2002; Berber Sardinha, 2004: 20-22; Sinclair, 2005; Tagnin, 2007a; Atkins e Rundell, 2008: 69). Com base nesses autores, e levando-se em conta os corpora produzidos para o presente trabalho, destaco alguns desses critérios:

- **língua:** os corpora podem conter textos escritos em apenas uma língua ou variedade lingüística (monolíngües), ou em mais de uma língua (bilíngües / multilíngües); podem ser ainda multivarietais (textos provenientes de variantes regionais ou nacionais distintas de uma mesma língua);
- **data de publicação:** de acordo com a data de publicação dos textos de um corpus, ele pode ser considerado

¹⁴⁵ *Um corpus pode ser descrito como uma extensa coleção de textos autênticos em formato eletrônico organizados de acordo com um conjunto específico de critérios.*

- sincrônico (recorta um período de tempo específico) ou diacrônico (contém textos representativos de diferentes períodos de tempo);
- histórico (contém textos de valor documental) ou contemporâneo (textos atuais);
- fechado (não há possibilidade ou intenção de se acrescentar mais textos após a conclusão da compilação) ou aberto;
 - corpus monitor (corpus aberto que é constantemente atualizado com a adição de novos textos de modo a refletir o estado atual de uma dada língua ou variedade lingüística);
- **modo:** um corpus pode conter textos escritos, transcrições de textos orais ou ambos;
- **conteúdo:** os textos do corpus podem ser representativos da língua geral (número balanceado de textos em gêneros e tipos textuais diversos), ou de qualquer parte da língua (especializados, regionais, etc.);
- **uso na pesquisa:** um corpus pode ser de estudo (aquele que se pretende analisar), ou de referência (em geral maior que o corpus de estudo, usado como termo de comparação);
- **autoria:** os textos podem ser escritos por falantes nativos, ou não-nativos (aprendizes de uma segunda língua, tradutores); podem ser de autoria individual, coletiva (artigos científicos, etc.) ou institucional (empresas, órgãos governamentais, etc.);
- **tamanho:** de acordo com o número de palavras de um corpus ele pode ser considerado pequeno (até 80 mil palavras); pequeno-médio (80 a 250 mil); médio (250 mil a 1 milhão); médio-grande (1 milhão a 10 milhões) e grande (acima de 10 milhões); é importante considerar o tamanho dos textos de um corpus, bem como a relação tamanho – autoria, para que a amostra não reflita idiosincrasias de autores ou predominância temática advindas de textos longos demais ou de mesma autoria;
- **nível de codificação:** os textos de um corpus podem ser acrescidos de diversas informações extratextuais e lingüísticas por meio da inserção de

cabeçalhos (vide Fig. 19) e etiquetas (morfo sintáticas, sintáticas, semânticas, discursivas ou pragmáticas – vide item 4.2.2.1 do Capítulo IV¹⁴⁶).

```
<Header>
  <title>
    <tit> Bolo inglês de Natal </tit>
    <filename> teb02_POB </filename>
    <subcorpus> receitas </subcorpus>
    <collection> tortas e bolos </collection>
  </title>
  <sourceText>
    <language> POB </language>
    <mode> Internet </mode>
    <status> caseira </status>
    <publisher> Revista Gula </publisher>
    <editor> Editora Peixe </editor>
    <pubPlace> http://www2.uol.com.br/gula/reportagens/121_simbolos_receitas.html </pubPlace>
    <accessDate> 12/fev/2003 </accessDate>
    <comments> a autora da receita viveu em Londres por muitos anos e, de volta ao Brasil,
    chegou a fabricar esse bolo para venda em parceria com o amigo e cozinheiro Charlô Whately </comments>
  </sourceText>
  <author>
    <name> Maria Helena Guimarães </name>
  </author>
</Header>
```

Figura 19: exemplo de cabeçalho preenchido

Os corpora multilíngües podem ser diferenciados ainda com relação à sua **organização interna**, em:

- **comparáveis** – compostos de dois ou mais subcorpora, cada um contendo textos originalmente escritos em uma dada língua ou variante lingüística, mas com características semelhantes, tais como tipologia textual, gênero, tamanho, data de publicação, área e subárea temáticas, etc. (p. ex.: 300 receitas de pão em português brasileiro e 300 receitas de pão em inglês americano).
- **paralelos** – compostos de textos originalmente escritos numa língua acompanhados de suas respectivas traduções para uma ou mais línguas
 - alinhados: sentenças do original e da(s) tradução(ões) podem ser vistas lado a lado ou uma abaixo da outra.

¹⁴⁶ para mais informações sobre etiquetagem e exemplos de uso vide Tagnin e Teixeira (2004b) e Teixeira (2007).

Embora haja uma recomendação explícita na literatura mais recente de que sejam usados textos integrais na composição dos corpora, há projetos que ainda trabalham com amostras, que tanto podem consistir de um fragmento de x palavras a partir do início do texto (p. ex.: o corpus Brown), quanto de uma porcentagem da obra total (p. ex.: o corpus COMPARA¹⁴⁷ inclui, no caso das obras literárias de seu acervo que ainda não caíram em domínio público, apenas 30% do texto) – essa escolha é justificável nos casos em que problemas com direitos autorais das obras impossibilitariam a compilação e disponibilização pública do corpus.

3.3.2 Representatividade

Sendo o corpus a fonte de todas as evidências lingüísticas analisadas em Lingüística de Corpus, a questão da representatividade deve ser discutida, obrigatoriamente, por um trabalho que adote uma metodologia como a que pretendo propor aqui, pois o sucesso na identificação de Unidades de Tradução Especializadas úteis ao tradutor vai depender do tamanho do corpus, da qualidade dos textos que o compõem, dos critérios empregados em sua compilação, de seu nível de codificação, entre outros fatores.

A questão da representatividade, como já foi comentado anteriormente, é muito controversa em L.C. Um dos motivos para isso é o fato de ter sido sempre um dos alvos de ataque preferidos dos gerativistas. Para esses pesquisadores, o corpus nunca será grande o bastante para representar a língua porque, em primeiro lugar, ele é meramente um conjunto de “exemplos”¹⁴⁸ de realizações lingüísticas (*performance*) representativas de uma estrutura universal, um esquema de possibilidades que todo falante nativo teria internalizado (*competência* – o verdadeiro objeto de estudo da lingüística para eles).

Além disso, estudos baseados em corpora analisariam a recorrência em dados lingüísticos que, na verdade, refletem os fatos do mundo – por exemplo: a frase "I live in New York" seria muito mais freqüente num corpus de língua geral do que "I live in Dayton, Ohio", pelo simples fato de que há muito mais pessoas morando New York do que em Dayton, Ohio. Berber Sardinha (2004: 32) contra-ataca esse famoso exemplo de Chomsky dizendo que, mesmo que se concluísse que a primeira frase é mais

¹⁴⁷ http://adamastor.linguatca.pt/COMPARA/compara_pt.html [acesso em 10/9/2008]

¹⁴⁸ vide citação de Rajagopalan à página 152

freqüente que a segunda num corpus (apenas uma suposição, que precisa ser verificada), ainda assim se teria aprendido algo sobre a língua – por exemplo, que “I live in CITY” é uma forma recorrente de se dizer, em língua inglesa, onde a pessoa mora, informação que pode ser da maior importância para um falante não nativo.

Vasilévski (2007) argumenta, nesse sentido, em favor de uma relativização da condição inalcançável da representatividade nos estudos com corpus:

(...) cabe a defesa de que se um *corpus* apropriado para pesquisa não representa a língua, então, ainda não há forma de fazê-lo, porque, seguramente, pesquisar a partir de dados inventados ou isolados constitui uma forma bem mais limitada, e nada representativa, de estudar a língua. Ainda, parece haver certo exagero quanto à falta de representatividade (...). Essa discussão pode culminar com a conclusão de que não há como estudar a língua, tendo em vista que, se uma densa e apropriada massa lingüística não a representa, o que mais pode representá-la? (Vasilévski, 2007: 53)

Vários autores propuseram metodologias para tentar contornar o problema da representatividade do corpus em L.C. (p. ex.: Stubbs, 1993; Sinclair, 1996; Leech, 1991; Sánchez e Cantos, 1997; Berber Sardinha, 1998 e 2002). Partindo do pressuposto defendido por Halliday (1991 e 1992) de que a língua é um sistema probabilístico, Berber Sardinha (2004: 22-30) procura responder às perguntas que ele mesmo coloca com relação à representatividade do corpus: “Representativo do quê e para quem?”.

Para responder à primeira pergunta, o autor defende que é preciso considerar a amostra relativamente à população que pretende representar – em qualquer cenário, quanto maior a amostra, maior a probabilidade de ser representativa de uma população cujo tamanho se desconhece. O tamanho, segundo ele, compreende três dimensões: i) número de palavras; ii) número de textos; iii) número de gêneros.

Uma das abordagens usadas em L.C. para determinar a representatividade do corpus é a Estatística (Biber, 1990 e 1993; Sánchez e Cantos, 1997). Berber Sardinha (2002) fez um estudo do tamanho médio dos corpora de acordo com o que se pretende estudar e a especialização do corpus. Para corpora especializados, como o compilado para a presente pesquisa, o autor calcula que todas as categorias gramaticais da língua inglesa (verbo, substantivo, adjetivo, advérbio, pronome e numeral) estão

representadas num corpus especializado com pouco mais de 91.000 palavras. No caso do estudo de itens lexicais específicos, dentre os exemplos citados pelo autor (*and, have, can, will, this, etc.*), o corpus que ora compilei só não seria suficiente para estudar a palavra “may” (que exigiria um corpus de cerca de 1,7 milhão de palavras – o corpus desta pesquisa contém cerca de 1,5 milhão em cada língua).

Para responder à pergunta “representativo para quem?” deve-se levar em conta, mais uma vez, que a representatividade, em termos estritos, não pode ser comprovada empiricamente quando não se conhece o tamanho da população que se procura representar. Portanto, a representatividade de um corpus, de acordo com Berber Sardinha, é atribuída pelos usuários do corpus (inclusive os próprios planejadores e compiladores do mesmo), de forma que:

O ônus de demonstrar a representatividade da amostra e de ser cuidadoso em relação à generalização dos seus achados para uma população inteira (um gênero ou língua inteiros, por exemplo) é dos autores. (Berber Sardinha, 2004: 25)

Por fim, a representatividade deve ser considerada também com relação à adequação do corpus aos propósitos a que se destina – um corpus considerado representativo de uma língua ou variedade lingüística pode mostrar-se inadequado à investigação de um aspecto lingüístico específico (p. ex: o corpus Brown, considerado representativo da língua inglesa, não é adequado para estudar o vocabulário da computação).

3.3.3 Padronização

A padronização da língua foi estudada por muitos autores em L.C., especialmente o fenômeno da colocação (p. ex.: Nattinger e DeCarrico, 1992; Eeg-Olofsson e Altenberg, 1996; Aijmer, 1996; Moon, 1998; Hudson, 1998; Biber, Conrad e Reppen, 1998), e recebe vários outros nomes na literatura: *prefabricated chunks, lexical bundles, idioms, multi-word units, n-grams, functionally complete units of meaning, etc.*

Sinclair (1991: 108-110) cunhou os termos “idiom principle” e “open-choice principle” para se referir aos dois extremos que governam a produção textual, seja ela escrita ou falada. Para os modelos teóricos que vêem a linguagem como uma estrutura

(inclusive o gerativista de Chomsky), a produção textual é puramente composicional, ou seja, se dá por meio de “esquemas *slot and filler*”, nos quais “as lacunas (*slots*) sintáticas podem ser preenchidas lexicalmente de qualquer modo, desde que o conjunto de lacunas seja estruturalmente plausível” (Berber Sardinha, 2004: 32). Por exemplo: suponha uma frase com a estrutura: artigo + substantivo + adjetivo + verbo transitivo direto + objeto direto (=artigo + substantivo + adjetivo). Seria possível “preencher as lacunas” dessas categorias de várias maneiras, algumas produziriam frases agramaticais, outras produziriam frases “gramaticalmente corretas”, mas nem todas fariam sentido em português brasileiro:

art.	subst.	adj.	v. trans. dir.	art.	subst.	prep.	subst.
A	menina	sapeca	pulou	a	cerca	de	arame
A	guria	ardilosa	transpôs	o	valado	de	alambre
As	garota	bonita	pintaram	os	muro	de	concreto
As	raparigas	azul	impetrou	um	bola	pelos	pedra
Um	velhos	recém-nascido	partiram	o	gato	com	terras
Uma	geladeira	coroca	lavou	os	pés	na	sopa
etc.	etc.	etc.	etc.	etc.	etc.	etc.	etc.

A frase “A menina sapeca pulou a cerca de arame” usada nesse exemplo foi inventada por nós. Por curiosidade, chequei sua freqüência na *web* digitando-a, entre aspas, na janela de busca do Google¹⁴⁹. O resultado obtido (em 27/4/2008) foi zero ocorrências. Se por um lado isso é uma prova de que a criatividade existe e de que é possível produzir linguagem usando o procedimento *slot and filler* (ou “open-choice principle”), também deixa evidente que o fato de eu ter criado uma frase gramaticalmente correta e semanticamente plausível não significa que ela seja usada com freqüência pelos falantes nativos do português – ainda que quanto mais longa a frase, maior a chance de ela ser única (ainda que plausível).

Outras duas frases gramaticalmente corretas que poderiam ser criadas usando palavras da mesma categoria gramatical e que respeitam as estruturas sintáticas da frase original (*slot and filler*) são *A *guria transpôs o valado de alambre* (correta gramaticalmente, possível semanticamente, mas muito improvável) e *Uma *geladeira coroca lavou os pés na sopa* (correta gramaticalmente mas improvável semanticamente). Ambas as frases não poderiam ser produzidas, no entanto, a partir

¹⁴⁹ <http://www.google.com>

do *idiom principle*, ainda que **As garota bonita pintaram os muro de concreto* seria, já que é gramaticalmente incorreta, mas semanticamente plausível e apresenta co-ocorrências prováveis de palavras, como “meninas bonitas” (23.200 oc. no Google¹⁵⁰) e “muro de concreto” (45.000 oc.).

A questão não é, portanto, negar a *possibilidade* de o *open-choice principle* ser usado para produzir frases corretas e/ou aceitáveis pelos falantes nativos de uma língua, mas sim atentar para a **probabilidade** de ocorrência de certas combinações de palavras. Isto porque, de acordo com o *idiom principle* descrito por Sinclair, a escolha de cada item lexical vai **pré-selecionar** um grupo mais restrito de outros itens com os quais **ocorre tipicamente**.

Se o artigo “Uma”, por exemplo, for selecionado na tabela do exemplo anterior, “velhos” e “raparigas” são eliminados da segunda coluna; se “geladeira” for a palavra escolhida entre as demais opções da segunda coluna, só nos resta “azul” na terceira coluna, e partir da quarta coluna a frase torna-se impossível, pois, ao que me parece, nenhum dos verbos lá arrolados co-ocorre tipicamente com geladeira.

O que a Lingüística de Corpus tem demonstrado claramente por meio da observação de grandes quantidades de textos escritos e orais disponíveis eletronicamente é que a linguagem é composta, em sua grande maioria, de frases semi-prontas, pré-fabricadas, que são escolhidas como um todo (“chunk” / “bundle”) pelo falante – o que Sinclair descreveu como *idiom principle*, mas que é muitas vezes referido como “fraseologia” ou “convencionalidade” por outros autores, e que estou chamando aqui de padronização.

Minha hipótese é a de que o mesmo princípio idiomático que governa a seleção lexical na língua geral opera também nas linguagens especializadas, mas os padrões mais prováveis são outros – ainda que os itens lexicais e as combinatórias morfossintáticas possam ser as mesmas. São esses, quando empregados apropriadamente pelo tradutor técnico, que darão ao texto a **naturalidade** almejada na tradução técnica (vide item 1.5 do Cap. I). Ainda no exemplo apresentado anteriormente, não somente a palavra “coroca” nos remete a “velha” (língua geral), mas “impetrar” nos remete a “ação”, e ao contexto específico dos textos jurídicos.

Quando se estuda a padronização de uma língua ou variedade lingüística procura-se identificar algum tipo de regularidade na co-ocorrência de palavras (lexical,

¹⁵⁰ pesquisas feitas em 12/09/2008

sintática, semântica, etc.). Para descrever tais padrões, o linguista de corpus, de acordo com Hoey (1997: 3), procura responder às seguintes questões:

- quais são os padrões lexicais de que a palavra faz parte?
- a palavra se associa regularmente a algum sentido específico?
- em que estruturas ela aparece?
- existe alguma correlação entre o uso ou o sentido da palavra e as estruturas de que participa?
- a palavra está associada a uma certa posição na organização textual?

Para descrever os padrões são usados, na literatura, três conceitos principais (Berber Sardinha, 2004: 40):

- a. **colocação**: associação entre itens lexicais (p. ex.: “presunto” co-ocorre com “*fatiado*” numa frequência maior do que com “*laminado*”);
- b. **coligação**: associação entre itens lexicais e gramaticais (p. ex.: “vá VERBO-GERÚNDIO *aos poucos*”: “*vá juntando / misturando / acrescentando aos poucos*”);
- c. **prosódia semântica**¹⁵¹: associação entre itens lexicais e conotação, que pode ser positiva, negativa ou neutra (p. ex.: “*gorduroso*” associa-se sempre a contextos negativos em culinária).

Vejamos, a seguir, algumas das metodologias e ferramentas de exploração de corpora mais usadas pela Lingüística de Corpus para descrever a padronização em corpora informatizados.

3.3.4 Ferramentas de exploração de corpora

O uso do computador para pesquisar os fenômenos da linguagem é um dos grandes diferenciais da L.C., pois permite a análise de um número muito maior de evidências lingüísticas – afinal, os computadores não se cansam, são rápidos e consistentes (não erram). As ferramentas computadorizadas de análise lingüística usadas pela L.C. para pesquisar os corpora são capazes de efetuar diversos cálculos

¹⁵¹ também chamada de preferência semântica e associação semântica (Hoey, 2003)

estatísticos comparativos, além de organizar os resultados das buscas de forma que os padrões lingüísticos sejam identificados com maior facilidade pelo pesquisador.

Três princípios abstratos norteiam as metodologias de estudo em L.C. (Berber Sardinha, 2004: 90): **ocorrência** (só se pode observar o que está no corpus); **recorrência** (para que se possa fazer generalizações, é preciso que um fenômeno ocorra com uma certa freqüência) e **co-ocorrência** (a ocorrência de um item deve ser relacionada com a ocorrência conjunta de outros fenômenos lingüísticos, sejam eles itens lexicais presentes no contexto imediato da palavra de busca, sejam características próprias do tipo e gênero textuais em que ocorre, ou mesmo fatores extralingüísticos). Algumas das ferramentas mais usadas por programas destinados a observar esse fenômenos em corpora são: listadores de palavras, concordanciadores (que podem apresentar também uma tabela de colocados e listas de clusters) e listadores de palavras-chave.

No que se segue, apresento cada uma dessas ferramentas usando como exemplo a versão 3 do programa WordSmith Tools (Scott, 1997)¹⁵², doravante WST, um dos *softwares* mais usados por pesquisadores da área¹⁵³ que trabalham dentro de um paradigma mais “informal” (isto é, baseado na análise de linhas de concordância¹⁵⁴), já que foi empregado em várias etapas desta pesquisa.

3.3.4.1 Listadores de palavras

Os listadores de palavras (ferramenta *WordList* do programa WST) apresentam uma lista de todas as palavras do corpus e sua respectiva freqüência; os dados podem ser consultados em três telas diferentes que apresentam: i) a lista das palavras por ordem decrescente de **freqüência** (F); ii) a lista das palavras por ordem **alfabética** (A) (que pode, inclusive, ser ordenada inversamente, ou seja, da última para a primeira letra da palavra – útil, para localizar, por exemplo, todas as palavras de um corpus de inglês terminadas em “-ing”); iii) uma tela com as **estatísticas** (S) do corpus, tais como o número de palavras distintas do corpus (*types*); o número total de palavras (*tokens*); a razão entre esses dois valores (razão forma/ocorrência – em inglês, *type/token ratio*);

¹⁵² uma versão gratuita de demonstração do programa pode ser baixada em: <http://www.lexically.net/wordsmith/> [acesso em 10/9/2008]

¹⁵³ para uma descrição completa das funcionalidades deste programa, vide capítulo 3 de Berber Sardinha (2004)

¹⁵⁴ Berber Sardinha (2004: 38)

o número de sentenças; de palavras com 1, 2, 3, etc. letras, entre outras. Tudo isso separadamente para cada subcorpus e para o corpus total (vide Fig. 20).

Figura 20: resultados da ferramenta WordList (listador de palavras)

É possível também elaborar listas de agrupamentos de palavras (2, 3... 8 palavras que ocorrem juntas, sempre na mesma ordem) – muito úteis na identificação de colocados, especialmente quando é possível compará-las com uma wordlist dupla, tripla, etc. de um corpus de língua geral, embora a maioria desses corpora seja grande demais para que sejam produzidas essas listas usando o WST¹⁵⁵. Para configurar o *Wordlist* para essa tarefa, clica-se sobre a opção *Settings* da barra de ferramentas, escolhe-se a opção *Min. & Max. frequencies*. Na tela que se abre, escolhe-se o número de palavras (*Clusters*), e deixa-se o box *Activated* marcado (vide Fig. 21). O próprio programa avisa que quanto maior o *cluster* escolhido, maior será o consumo de memória RAM. Em seguida, procede-se à compilação da lista de palavras normalmente. Um exemplo das telas de resultado obtidas pode ser visto na figura 22.

¹⁵⁵ O programa WST, versão 3, só aceitou fazer a wordlist do *corpus* de português Lácio-Ref em meu computador com 2 palavras – a lista com 3 palavras nunca é concluída e apresenta erro, travando o computador ao final.

Figura 21: configuração para compilar wordlists com mais de 1 palavra

N	1	2
Text File	OVERALL	IR_H-1.TXT
Bytes	655.520	575.634
Tokens	117.157	103.111
Types	9.200	36.900
Type/Token Ratio	7,85	35,79
Standardised Type/Token	88,59	88,87
Ave. Word Length	4,28	4,27
Sentences	2.961	2.588
Sent.length	26,07	27,30
sd. Sent. Length	15,62	15,56
Paragraphs	1.488	1.109
Para. length	78,73	92,98
sd. Para. length	117,15	130,51
Headings	0	0
Heading length		
sd. Heading length		
1-letter words	20.270	18.038
2-letter words	20.002	17.324
3-letter words	12.936	11.740
4-letter words	11.083	9.766
5-letter words	14.598	12.637
6-letter words	13.250	11.828
7-letter words	10.033	8.615
8-letter words	7.053	6.229
9-letter words	3.981	3.513
10-letter words	2.028	1.819
11-letter words	873	723
12-letter words	608	502
13-letter words	172	150

N	Word	Freq.	%	Lemmas
1	A # CM	2		
2	A # COLHERES	6		
3	A # LITRO	2		
4	A # MINUTOS	17		
5	A # PORÇÕES	3		
6	A ## DO	3		
7	A ## XICARA	3		
8	A #FC ALTO	29		
9	A #FC MÉDIO	133		
10	A #FC MUITO	2		
11	A #FC POR	8		
12	A ABOBRINHA E	7		
13	A ABOBRINHA REFOGADA	2		
14	A ACIDEZ COM	18		
15	A ÁGUA AOS	5		
16	A ÁGUA COM	3		
17	A ÁGUA DE	4		
18	A ÁGUA DO	2		
19	A ÁGUA E	22		
20	A ÁGUA FERVENTE	6		
21	A ÁGUA FERVER	2		
22	A ÁGUA MORN	3		
23	A ÁGUA NUMA	5		
24	A AJUDA DE	6		
25	A ALCAPARRA E	2		
26	A AMÊNDOA E	7		
27	A AMÊNDOA NUMA	5		
28	A ANCHOVA E	2		

N	Word	Freq.	%
38	UM POUCO DE	83	0,07
39	AJUSTE O SAL	81	0,07
40	E # MINUTOS	81	0,07
41	ATÉ OBTER UMA	80	0,07
42	DE CREME DE	80	0,07
43	FOLHAS DE SALSINHA	80	0,07
44	HORA DE SERVIR	78	0,07
45	O SUCO DE	77	0,07
46	XÍCARA DE FOLHAS	76	0,06
47	EM CUBINHOS #	75	0,06
48	FILME PLÁSTICO E	75	0,06
49	DE AZEITE DE	73	0,06
50	PELO MENOS #	73	0,06
51	NA GELADEIRA POR	72	0,06
52	ATÉ # HORAS	71	0,06
53	DE MANTEIGA #	71	0,06
54	FIO DE AZEITE	70	0,06
55	HORA E #	69	0,06
56	COM FILME PLÁSTICO	67	0,06
57	CHÁ DE SAL	66	0,06

Figura 22: Resultados do Wordlist com *clusters* de 3 palavras

3.3.4.2 Listadores de palavras-chave

Palavras-chave, em Lingüística de Corpus, são aquelas cuja frequência relativa no corpus de estudo é estatisticamente significante se comparada à frequência relativa

da mesma palavra num corpus de referência (Berber Sardinha, 2004: 96). O corpus de referência é aquele usado como termo de comparação. No caso do presente trabalho, por exemplo, o corpus de referência usado para o português foi o Banco de Português – corpus monitor de língua geral mantido pelo LAEL da PUC-SP que, em 2003, contava com cerca de 240 milhões de palavras¹⁵⁶. Em geral, espera-se que o corpus de referência seja cerca de 5 vezes maior que o corpus de estudo (quanto menor, menos palavras-chave serão identificadas pelo programa), não contenha os textos do corpus de estudo e, de preferência, seja composto de textos que vão colocar em evidência as características peculiares do corpus que se pretende estudar. Por exemplo: se o objetivo é observar as características do texto científico, é importante que o corpus de referência não contenha uma quantidade desproporcional desse tipo de texto, ou os itens lexicais que lhe são peculiares serão anulados durante a comparação.

A tela de resultados da ferramenta *KeyWords* do WST (vide Fig. 23) mostra, em sete colunas: as palavras-chave do corpus de estudo por ordem decrescente de chavidade (*keyness*); a frequência dessas palavras no corpus de estudo; a porcentagem que essa frequência representa com relação ao corpus de estudo total; a frequência dessas palavras no corpus de referência; a porcentagem que essa frequência representa com relação ao corpus de referência total; a chavidade, ou seja, o resultado da estatística de comparação (pode-se usar tanto o *log likelihood* quanto o *chi-square*) e o *p*, valor de significância estatística atingido pelo resultado da chavidade, ou seja, a probabilidade de aquela palavra ter sido erroneamente identificada como chave pelo cálculo estatístico¹⁵⁷. Por exemplo: um *p* de 0,0001 significa 1 chance em mil de aquela palavra ter sido apontada como chave por acaso.

Quando se observa, por exemplo, as palavras 4 e 5 da figura 23, “colheres” e “preparo”, é possível perceber que a segunda, ainda que tenha um número maior de ocorrências que a primeira no corpus de estudo, tem uma chavidade menor nesse corpus (89.332,3), pois ocorre também com uma frequência bem maior que “colheres” no corpus de referência – ou seja, o item lexical “colheres” parece ser mais restrito à área de especialidade representada pelo corpus do que “preparo”, ou qualquer outra palavra nas posições seguintes da lista.

¹⁵⁶ parte do *corpus* pode ser consultado em: <http://www2.lael.pucsp.br/corpora/bp/conc/index.html>

¹⁵⁷ para mais detalhes sobre o cálculo da chavidade, da estatística de associação e do *p*, vide, por exemplo, o *Help* do programa WST e Rocha (2007).

N	WORD	FREQ	TESEPOTT LST %	FREQ	BANCOP~1 LST %	KEYNESS	P
1	SOPA	15.115	0,99	3.192		135.226,5	0,000000
2	SAL	12.844	0,84	3.621		111.950,8	0,000000
3	CHÁ	11.226	0,74	4.202		94.943,0	0,000000
4	COLHERES	9.705	0,64	945		91.274,1	0,000000
5	PREPARO	10.209	0,67	2.758		89.332,3	0,000000
6	XICARA	8.994	0,59	1.004		83.979,8	0,000000
7	INGREDIENTES	9.134	0,60	3.728		76.457,3	0,000000
8	COLHER	8.862	0,58	3.405		74.710,3	0,000000
9	COLOQUE	8.216	0,54	2.638		70.650,1	0,000000
10	MANTEIGA	7.015	0,46	1.958		61.176,9	0,000000
11	MINUTOS	10.549	0,69	35.985	0,02	56.795,9	0,000000
12	MISTURE	5.640	0,37	643		52.588,9	0,000000
13	RETIRE	5.594	0,37	908		51.020,8	0,000000
14	DEIXE	6.588	0,43	4.953		50.567,2	0,000000
15	PANELA	5.760	0,38	1.518		50.503,4	0,000000
16	FOGO	8.179	0,54	19.577		48.881,7	0,000000
17	FORNO	5.739	0,38	2.077		48.702,5	0,000000
18	FARINHA	5.910	0,39	3.091		47.906,8	0,000000
19	JUNTE	4.999	0,33	680		46.127,4	0,000000
20	AZEITE	5.237	0,34	1.392		45.879,2	0,000000
21	ÁGUA	9.221	0,61	41.810	0,02	45.098,7	0,000000
22	MOLHO	5.766	0,38	3.973		44.880,4	0,000000
23	LEITE	7.529	0,50	20.525		43.373,6	0,000000
24	TEMPERATURA	6.122	0,40	7.386		43.068,9	0,000000
25	CEBOLA	5.046	0,33	1.794		42.904,6	0,000000
26	AÇÚCAR	6.422	0,42	9.931		42.817,5	0,000000
27	G	6.705	0,44	13.006		42.341,7	0,000000
28	CREME	5.063	0,33	2.551		41.246,7	0,000000
29	LEVE	6.140	0,40	9.425		41.005,2	0,000000
30	ACRESCENTE	4.461	0,29	734		40.645,1	0,000000
31	MASSA	6.414	0,42	12.975		40.073,8	0,000000
32	DE	125.837	8,27	10.729.884	4,66	36.478,4	0,000000
33	PIMENTA-DO-REINO	3.557	0,23	0		35.773,0	0,000000
34	ALHO	4.236	0,28	1.731		35.439,3	0,000000
35	SIRVA	3.870	0,25	1.131		33.589,6	0,000000
36	DIFICULDADE	5.640	0,37	14.143		33.266,4	0,000000
37	MODO	6.464	0,43	26.318	0,01	32.815,5	0,000000
38	QUENTE	4.765	0,31	6.554		32.604,2	0,000000

Figura 23: resultado da ferramenta KeyWords (listador de palavras-chave)

3.3.4.3 Concordanciadores

Esse tipo de ferramenta (*Concord* no WST) organiza na tela todos os contextos em que uma determinada palavra (*nódu*lo ou *palavra de busca*), escolhida pelo usuário, ocorre. As linhas ficam dispostas uma abaixo da outra e a palavra pesquisada aparece centralizada em todas elas, facilitando a identificação de padrões lingüísticos (vide Fig. 24). É possível também, no programa WST, reordenar alfabeticamente os resultados de várias maneiras para identificar diferentes tipos de padrões como, por exemplo, os formados pelas palavras que ocorrem de um lado e de outro da palavra de busca, que são chamadas de *colocados* em L.C. (do inglês *collocates*) – vide Fig. 25, com as linhas de concordância ordenadas pelos colocados à direita do nódulo, e Fig. 26, em que as linhas foram ordenadas pelos colocados à esquerda do nódulo.

Concord - [PIMENTÃO: 1064 entries (sort: 5L,5L)]									
Concordance					Set	Ta	Word No.	File	%
1032	a	1 colher(sopa) de pimentão verde picado	1 colher(sopa) de pimentão vermelho picado	2 colheres(sopa) de suco de limão			94.982	especp-1.txt	95
1033	imenta-do-reino	a gosto 1 pimentão verde médio em pedaços	1 pimentão vermelho médio em pedaços	1 cebola grande em ro			179.715	1'terrat-1.txt	20
1034	lheres (sopa)	de óleo 1 pimentão verde em pedaços médios	1 pimentão vermelho em pedaços médios	1 cebola roxa grande			416.691	1'terrat-1.txt	45
1035	ODELAS;	1 PIMENTÃO VERDE CORTADO EM RODELAS;	1 PIMENTÃO VERMELHO CORTADO EM RODELAS;	1 PIMEN			15.109	palcpo-1.txt	62
1036	ogumelos frescos	fatiados 40 g de milho verde cozido	40 g de pimentão vermelho em pequenos cubos	40 g de ervilhas fresc			46.559	basigd-1.txt	88
1037	es	picado 1/2 xícara de cebolinha verde picada	1/2 xícara de pimentão vermelho picado	3 colheres (sopa) de azeite	2 co		18.825	claudi-1.txt	15
1038	s de sopa	de óleo 8 espigas de milho verde	1 cebola picada 1 pimentão verde picado	1 e 1/2 xícara de creme de leite	1 colhe		166.324	1'terrat-1.txt	18
1039	a (chá)	de arroz cru 1 pimentão verde em pequenos pedaços	1 pimentão vermelho em pequenos pedaços	1 cebola média rala			152.031	1'terrat-1.txt	17
1040	ementes	picado 1/2 pimentão verde sem sementes picado	1/2 pimentão amarelo sem sementes picado	1 maço médio de sal			564.532	1'terrat-1.txt	61
1041	lete	1 abobrinha pequena e bem verde, em rodinhas finas	1/2 pimentão vermelho em cubinhos	2 tomates bem vermelhos, s			95.067	helopb-1.txt	94
1042	nela,	ponha o catado, os temperos verdes picados, a cebola, o	pimentão e os tomates em rodela por cima e o dendê. Leve a				71.681	especp-1.txt	72
1043	decore	com azeitonas verdes, camarões reservados, tiras de	pimentão e rodela de cebola. Leve ao forno pré-aquecido e m				187.289	1'terrat-1.txt	20
1044	es	em cubos 1/2 pimentão vermelho sem sementes picado	1/2 pimentão verde sem sementes picado	1/2 pimentão amarelo s			564.526	1'terrat-1.txt	61
1045	tomate	sem semente 1 pimentão vermelho cortado em tiras	1 pimentão verde cortado em tiras	1 cebola grande picadinha 3 c			297.555	1'terrat-1.txt	32
1046	ELAS;	1 PIMENTÃO VERMELHO CORTADO EM RODELAS;	1 PIMENTÃO AMARELO CORTADO EM RODELAS;	2 TOMATE			15.115	palcpo-1.txt	62
1047	dia	cortada em gomos 1 pimentão vermelho médio em tiras	1 pimentão amarelo médio em tiras	2 xícaras (chá) de alfafe	2 x		436.071	1'terrat-1.txt	47
1048	redientes:	Recheio 1 pimentão vermelho 1 pimentão verde	1 pimentão amarelo 500 g de batata em cubinhos	200 g de peito			282.143	1'terrat-1.txt	31
1049	tomate	sem semente 1 pimentão vermelho cortado em tiras	1 pimentão verde cortado em tiras	1 cebola grande picadinha 3 c			379.791	1'terrat-1.txt	41
1050	de alho	amassado 1 pimentão vermelho cortado em rodela	1 pimentão verde cortado em rodela	1 pimentão amarelo cortad			750.827	1'terrat-1.txt	82
1051	s pelo	comprimento 1 pimentão vermelho cortado em tiras	1 pimentão amarelo cortado em tiras	Molho 1/2 xícara de vin			67.111	claudi-1.txt	52
1052	e ricota	3 cenouras 1 pimentão vermelho 1 pimentão verde	1 pimentão amarelo 1 talo de salsão	1 maço pequeno de raban			912.651	1'terrat-1.txt	99
1053	: Difícil	Chutney de pimentão vermelho	Ingredientes: 1/2 kg de pimentão vermelho	1/2 kg de tomate maduros	1/2 kg de cebolas		488.969	1'terrat-1.txt	53
1054	gosto	4 berinjelas médias 1 pimentão vermelho em tiras finas	1 pimentão verde em tiras finas	Modo de Preparo: Descasque			396.776	1'terrat-1.txt	43
1055	e pião	de tomate 1 pimentão vermelho pequeno em rodela	1 pimentão verde pequeno em rodela	2 dentes de alho 1 colher			872.798	1'terrat-1.txt	95
1056	la	cortada em tiras 1/2 pimentão vermelho cortado em tiras	1/2 pimentão verde cortado em tiras	1/2 pimentão amarelo cortado			608.870	1'terrat-1.txt	66
1057	tos,	virando o frango e a linguiça de vez em quando. Adicione o	pimentão, a cebola e o alho. Junte o açafrão, a folha de louro e				763.606	1'terrat-1.txt	83
1058	lata	de creme de leite; 1 vidro de champignon; Azeite; Tomate;	Pimentão; Cebola; Cheiro-verde; 2 1/2 xícaras de leite; 3 colher				53.180	especp-1.txt	54
1059	os 1/4	de xícara (chá) de vinagre pimenta-do-reino a gosto	1/2 pimentão verde picado	600 g de suco de tomate	2 colheres de		754.450	1'terrat-1.txt	82
1060	Modo de fazer:	Prepare o vinagrete: no liquidificador, coloque o	pimentão, o alho, o pepino, o tomate, o ovo, o vinagre, o suco d				91.264	especp-1.txt	91
1061	o de manjeriça	1 litro de vinagre de vinho branco 1 berinjela	1 pimentão vermelho 1 erva-doce pequena	2 abobrinhas azeite			126.699	1'terrat-1.txt	14
1062	branco	seco 30 ml de azeite extra virgem 4g de páprica doce	pimentão 8g de tinta de lula 50g de amêndoas torradas e pica				26.197	1'jpant-1.txt	64
1063	rado	Pimentão Rico em vitamina C, betacaroteno e cálcio,	pimentão faz bem ao coração e ao estômago. Também reforç				981	basigd-1.txt	2
1064	500g	de camarões médios limpos 1 xícara de cebola picada	1 pimentão verde pequeno picado	6 tomates picados sem seme			393.367	1'terrat-1.txt	43

Figura 24: resultado de busca para “pimentão” no concordanciador (Concord)

Concord - [PIMENTA: 6263 entries (sort: 1R,2R,3R)]											
File View Settings Window Help											
Concordance											
								Set	Ta	Word No.	File
1	a	de ovo crua	Modo de preparo	Tempere a carne com o sal, a pimenta, a salsinha, a cebolinha, o louro, a	cebola e o vinho. Deix			28.399	direct-1	txt	
2	entralizado	Colocar o molho sobre o peixe	Prepara um molho de	pimenta a base	de shoyo light, mais pimenta	Dedo de Moça	corta	35.683	1'jpant-1	txt	
3	com cerveja	Ingredientes: 2 cebolas médias picadas	sal a gosto	pimenta a gosto	1 folha de louro	2 kg de lagarto	2 xícaras (chá) d	920.375	1'terrat-1	txt	1
4	Ingredientes	Frango 1/2 kg de filés de frango	Sal, ervas finas,	pimenta a gosto	1 colher (sopa) de mostarda	Suco de limão	1 c	8.314	direct-1	txt	
5	laranja	Ingredientes: 1 kg de contrafilé sem gordura	sal a gosto	pimenta a gosto	1 dente de alho	1 colher (chá) de vinagre	1 colh	593.067	1'terrat-1	txt	
6	unte o suco de limão,	o creme de leite, a salsinha, acerte o sal e a	pimenta (bata no liquidificador se quiser uma pasta mais lisa), cu					73.455	elopb-1	txt	
7	maduro, sem pele e sem semente em gomos	1/4 colher (café) de	pimenta branca	1 colher (chá) de maisena	1 ramo de cheiro-ver			361.121	1'terrat-1	txt	
8	é) de	sal 1 peito de frango grande em 2 filés grossos	2 pitadas de	pimenta branca	1 colher (sopa) de farinha de trigo	1 ovo	1/2 xica	517.880	1'terrat-1	txt	
9	avirgem (para decoração)	1 buquê de manjeriça verde-roxo	sal e	pimenta branca	moída a gosto	ovas de salmão para decorar	(opci	606.687	1'terrat-1	txt	
10	o 500 g	de tomate sem pele e sem semente	500 g de talo de aipo	pimenta branca	moída na hora, a gosto	50 g de amêndoas picada		212.291	1'terrat-1	txt	
11	ol. (sobremesa)	de azeite balsâmico 4 pitadas de sal	pitadas de	pimenta branca	moída na hora	4 col. (sopa) de azeite de oliva ext		9.643	asigd-1	txt	
12	á) de	cebola em pó 1 colher (chá) de alho em pó	1 colher (chá) de	pimenta caiena	1 colher (chá) de orégano	1 colher (chá) de tomil		634.882	1'terrat-1	txt	
13	1/2 kg	de lombo de porco 2 colheres (chá) de sal	1 colher (chá) pimenta	caiena	1 colher (chá) de pimenta-do-reino	moída 1 colhe		99.249	1'terrat-1	txt	
14	tes 1	colher (sopa) de alho socado	sal a gosto 1/2 colher (chá) de	pimenta caiena	4 colheres (sopa) de vinagre	1 xícara (sopa) de		728.433	1'terrat-1	txt	
15	tes 1	colher (sopa) de alho socado	sal a gosto 1/2 colher (chá) de	pimenta caiena	4 colheres (sopa) de vinagre	1 xícara (sopa) de		799.131	1'terrat-1	txt	
16	a-doce	2 cenouras 1 colher (sopa) extrato de tomate	sal a gosto	pimenta caiena	a gosto 1 e 1/2 litro de água	1/2 copo de vinho br		148.877	1'terrat-1	txt	
17	adinhos	2 colheres (chá) de gengibre picadinho	1 colher (chá) de	pimenta-calabresa	1 colher (chá) de cominho em pó	1 colher (ch		77.007	elopb-1	txt	
18	de alho	picados 1 colher de sopa de orégano	1 colher de café de	pimenta calabresa	1 colher de chá de sementes de erva-doce	5		781.511	1'terrat-1	txt	
19	óleo	2 colheres (sopa) de manjeriça picado	1 colher (chá) de	pimenta-calabresa	1 kg de tomate sem pele e sem sementes pic			4.741	claudi-1	txt	
20	heres (sopa)	de azeite 1 colher (chá) de sal	1/2 colher (chá) de	pimenta calabresa	1 kg de vôngoles com as conchas, limpos e l			50.075	claudi-1	txt	
21	até o	frango ficar cozido e macio. Sirva o frango acompanhado de	pimenta cambuci	frita. Fonte: Revista Água na Boca	Categoria			912.620	1'terrat-1	txt	
22	(chá) de salsinha	picada 1 xícara (chá) de coalhada	sal a gosto	pimenta da jamaica	para decorar	Modo de Preparo: Lave o pepin		621.762	1'terrat-1	txt	
23	ta-do-reino	branca em pó, gengibre em pó, noz moscada ralada e	pimenta da Jamaica	ralada. Categoria: Lanches	Cozinha: Árab			22.274	1'terrat-1	txt	
24	a) rasa	de sal 1/2 col. (café) de pimenta branca	moída 6 grãos de	pimenta-da-jamaica	1 col. (chá) de pimenta-rosa	2 cravos-da-in		25.529	asigd-1	txt	
25	tado 1/2	kg de tomate em conserva sem pele	2 colheres (chá) de	pimenta-da-jamaica	2 colheres (sopa) de manjerona	picada 2 co		70.373	1'terrat-1	txt	
26	tado 1/2	kg de tomate em conserva sem pele	2 colheres (chá) de	pimenta-da-jamaica	2 colheres (sopa) de manjerona	picada 2 col		571.896	1'terrat-1	txt	
27	manjeriça	seco 1 colher de café de orégano	1 colher de café de	pimenta-de-caiena	1 colher de café de páprica picante	1 xícara d		886.299	1'terrat-1	txt	
28	(sopa)	de cherry 2 colheres (sopa) de molho de soja	1 pitada de	pimenta-de-caiena	125 ml de óleo	Modo de preparo	Descasqu	39.594	direct-1	txt	
29	opa)	de uvas passas brancas 1 e 1/2 xícara (chá) de arroz	sal e	pimenta-de-cheiro	a gosto	Modo de Preparo: Lave os shiitakes,		383.537	1'terrat-1	txt	
30	a, de	salsa picada; 2 colheres, das de sopa, de cebolinha picada;	Pimenta-de-cheiro	a gosto; 1 xícara de chá de óleo;	250 ml de leit			54.864	specp-1	txt	
31	ola	grande cortada em rodela finas	4 dentes de alho picados	1 pimenta-de-cheiro	amassada	Sal a gosto	Preparo	1. Deixe o	68.305	claudi-1	txt
32	pimenta	Ingredientes: geleia de pimenta com abacaxi	150 g de	pimenta dedo-de-moça	1 abacaxi médio	1 cebola média	5 colher	599.516	1'terrat-1	txt	
33	ebola	média 2 dentes de alho	2 tomates	1/2 pimentão verde	1/2 pimenta	dedo-de-moca	1 colher de soa	de coentro picado	2 col	61.460	1'terrat-1

Figura 25: resultado de “pimenta” ordenado pelos colocados à direita

N	Concordance	Set	Ta	Word No	File
1	damente o azeite de oliva. Corrigir o ponto de sal e acrescentar a pimenta-do-reino. Sirvir com pão ou torradas. Categoria: Acomp			277.803	1'terrat-1.txt
2	do e misture bem. Quando o queijo estiver derretido, acrescente a pimenta e, se necessário, a água reservada. Misture novamente e			646.118	1'terrat-1.txt
3	re, vinho e sal. Misture bem para que o sal dissolva. Acrescente a pimenta-do-reino, o orégano e o alho amassado. Vá acrescentan			112.145	1'terrat-1.txt
4	bem até obter uma mistura homogênea. Acerte o sal e adicione a pimenta. Transfira metade da mistura para uma folha de papel-m			879.717	1'terrat-1.txt
5	ebola picada. Refogue até que a cebola esteja macia e adicione a pimenta dedo-de-moça, refogue por 30 segundos			184.667	1'terrat-1.txt
6	e caldo de carne ou de galinha até cobrir o arroz. Tempere com a pimenta-do-reino branca e deixe cozinhar normalmente. Após uns			46.712	especp-1.txt
7	to isso, lave a carne, seque com toalha de papel e tempere com a pimenta-do-reino. Coloque uma frigideira de ferro no fogo até ficar			558.870	1'terrat-1.txt
8	, sem parar de mexer, espere ferver e engrossar. Acerte o sal e a pimenta e, se for preciso, ajuste a acidez com 1 pitada de açúcar,			49.801	1'helobp-1.txt
9	um batedor manual até obter um molho cremoso. Acerte o sal e a pimenta . Volte o peixe para a frigideira, aqueça bem e distribua e			883.459	1'terrat-1.txt
10	s. Regar com conhaque e juntar aos camarões. Retificar o sal e a pimenta , se necessário. Ferver por mais 5 minutos. Colocar o cre			5.900	1'terrat-1.txt
11	Modo de Preparo: Misture o óleo com o suco de limão, o sal e a pimenta e tempere o peixe. Cozinhe as folhas de espinafre em po			432.013	1'terrat-1.txt
12	duro, cozinhar mais um pouco em fogo brando. Retificar o sal e a pimenta , se necessário. Servir acompanhado de arroz. Categori			489.274	1'terrat-1.txt
13	spas das cascas dos cítricos, o sal grosso e o açúcar. Polvilhe a pimenta e deixe marinar por 3 horas na geladeira. •Retire o exc			3.892	1'grgula.txt
14	aras. Faça 8 medalhões com cerca de 3 cm de altura e polvilhe a pimenta-do-reino. Coloque o azeite restante numa frigideira grand			603.746	1'terrat-1.txt
15	o chuchu nos pratos, coloque por cima um ovo pochê e polvilhe a pimenta-do-reino. Sirva com salada de alface e pedaços de polpa			392.428	1'terrat-1.txt
16	eiro corte em rodela. Reserve. Regue com o azeite, salpique a pimenta-do-reino, decore com rodela de limão e pimenta-rosa. C			48.323	1'basigd-1.txt
17	ndo do prato (como se fosse um carpaccio de carne). Salpique a pimenta do reino (usar pouco mesmo) Por cima, novamente, um			12.262	1'jpanttt-1.txt
18	he o sal e retire do fogo. Distribua os bifes nos pratos e salpique a pimenta. Sirva com salteado de pimentão com ervas de proven			172.116	1'terrat-1.txt
19	a frigideira e cozinhe até obter um molho aveludado. Decore com pimenta rosa. Fonte: Revista Água na Boca Categoria: Molhos			188.255	1'terrat-1.txt
20	ira por 2 horas, ou até que fique firme. Desenforme e decore com pimenta-rosa. Fonte: Revista Água na Boca Categoria: Acomp			868.540	1'terrat-1.txt
21	massa nos pratos. Cubra com o molho preparado e decore com pimenta-dedo-de-moça. Sirva em seguida. Fonte: Revista Menu			62.146	1'terrat-1.txt
22	certe o sal e cozinhe por mais 5 minutos. Se preferir, decore com pimenta e sirva a seguir. Fonte: Revista Água na Boca Catego			903.670	1'terrat-1.txt
23	em ponto de fio. Quando a massa estiver cremosa, tempere com pimenta e com noz-moscada, se quiser. Arrume as azeitonas e			45.252	1'cdrect-1.txt
24	fogue em fogo baixo e acrescente a carne desfiada, tempere com pimenta-do-reino e o coentro. Prepare o purê de mandioca, pass			60.708	1'terrat-1.txt
25	reparo: Abra os filés de peito de frango bem finos e tempere com pimenta-do-reino (sem sal). Disponha 1 fatia de presunto cru e a			612.935	1'terrat-1.txt
26	s 1 1/2 colher de chá de sal 1 colher de chá de molho forte de pimenta (opcional) 1/2 colher de chá de pimenta do reino 1 co			67.276	1'especp-1.txt
27	cada um ao meio sem separar as partes. Coloque molho forte de pimenta dentro e camarões cozidos (ou vatapá, como se faz na			67.405	1'especp-1.txt
28	xicara de vinagre branco 1 cenoura 1 cebola graúda 10 grãos de pimenta-do-reino 1 folha de louro 4 talos de salsinha 1 litro de a			821.294	1'terrat-1.txt
29	salsão picado. Junte 1 xícara (chá) de vinho branco, 12 grãos de pimenta-do-reino, 2 folhas de louro, suco de 1/2 limão e 2,5 litros			432.727	1'terrat-1.txt
30	po com folhas em pedaços grandes 1 bouquet garni 12 grãos de pimenta preta ou branca Modo de preparo Numa panela, colo			114.860	1'cdrect-1.txt
31	de louro 1/2 colher (chá) de tomilho seco sal a gosto 12 grãos de pimenta-do-reino 3 ramos de salsinha Modo de Preparo: Coloq			295.147	1'terrat-1.txt
32	chup 1/3 de xícara de molho de ostra 1 colher (sopa) de molho de pimenta-vermelha e sal para o molhinho de mexerica 1 colher (so			82.967	1'helobp-1.txt
33	s de sopa de sal 3 folhas de louro 1 colher de sopa de molho de pimenta vermelha 2 limões 1/4 de xícara de tempero do Luiz 100			137.911	1'terrat-1.txt

Figura 26: resultado de “pimenta” ordenado pelos colocados à esquerda

É possível também expandir o contexto mostrado para ler um trecho maior do texto de origem ou mesmo acessá-lo integralmente, a partir da barra de ferramentas. Nas colunas da direita, é possível ver, entre outras coisas, o arquivo de onde aquele contexto provém. A ferramenta apresenta uma série de outros recursos que não cabe detalhar aqui (vide Fig. 27). Apresento, a seguir, dois que estão diretamente relacionados com esta pesquisa: a tabela de colocados e a lista de *clusters*.

Figura 27: funções acessadas pela barra de ferramentas do Concord

3.3.4.4 Tabela de colocados

Uma tabela de colocados mostra as palavras mais recorrentes à esquerda e à direita de um nóculo (geralmente 5 para cada lado) em forma de tabela. No programa WST, a tabela de colocados pode ser acessada pelo nono ícone da barra de ferramentas da tela de resultados do *Concord* (vide Fig. 28). Para obter o resultado exibido na figura 28, por exemplo, foi feita inicialmente uma busca por “picad*” no corpus de português (o asterisco, ou caractere coringa, permite recuperar todas as formas que contêm a seqüência de letras que está antes ou depois do mesmo – neste caso: picado, picada, picadas, picadinhos, etc.). Em seguida, a tabela de colocados foi acessada clicando-se o ícone apropriado.

N	WORD	TOTAL	LEFT	RIGHT	L5	L4	L3	L2	L1	*	R1	R2	R3	R4	R5
1	DE	10796	5686	5110	710	970	943	3061	2	0	17	186	1486	2023	1398
2	PICADA	3706	232	216	153	47	26	6	0	3258	3	17	40	81	75
3	SOPA	2647	1494	1153	133	244	1099	10	8	0	0	0	633	445	75
4	E	2346	1096	1250	146	184	375	166	225	0	475	222	190	170	193
5	PICADO	2250	162	161	74	51	27	10	0	1927	1	3	16	42	99
6	CHÁ	1723	838	885	77	196	555	5	5	0	0	1	479	329	76
7	XÍCARA	1581	824	757	199	432	192	0	1	0	0	396	225	94	42
8	CEBOLA	1561	1368	192	39	59	57	459	754	1	0	85	56	16	35
9	A	1480	702	778	139	73	99	389	2	0	97	310	110	118	143
10	COLHERES	1435	672	763	221	442	8	1	0	0	0	498	167	38	60
11	COLHER	1362	651	711	210	422	19	0	0	0	0	464	194	28	25
12	PICADOS	1209	103	128	35	44	22	2	0	978	3	2	13	34	76
13	PICADAS	1201	93	67	42	33	14	4	0	1041	0	2	11	19	35
14	ALHO	1193	840	353	82	73	39	86	560	0	0	9	16	189	139
15	SALSINHA	1013	881	132	15	28	28	39	771	0	14	9	18	16	75
16	O	993	538	455	140	58	38	297	5	0	3	149	118	89	96
17	G	943	501	442	180	126	190	3	2	0	4	280	95	25	38
18	SAL	707	96	611	31	40	15	8	2	0	149	169	79	80	134
19	EM	643	176	467	64	50	8	52	2	0	205	15	66	65	116
20	PIMENTA	625	166	459	29	88	20	19	10	0	9	47	197	136	70
21	SEM	612	397	215	123	29	46	199	0	0	39	17	34	71	54
22	GOSTO	546	115	431	35	20	17	41	2	0	0	96	206	58	71
23	CEBOLINHA	540	435	105	7	9	14	205	200	0	2	13	6	16	68
24	DENTES	528	309	219	20	24	265	0	0	0	0	125	81	5	8
25	BEM	509	364	145	25	13	6	1	319	0	25	2	72	22	24
26	COM	503	285	218	55	43	105	82	0	0	41	15	94	18	50
27	DO	447	97	350	37	12	14	34	0	0	1	25	40	165	119
28	SALSA	443	404	39	5	14	7	21	357	0	4	3	11	10	11
29	VERDE	437	356	81	16	9	8	26	297	0	0	4	44	16	17
30	MODO	430	0	430	0	0	0	0	0	0	32	184	54	109	51
31	COL	428	227	201	40	186	1	0	0	0	0	28	153	12	8
32	TOMATES	395	241	154	20	39	31	59	92	0	0	78	43	21	12

Figura 28: tabela de colocados de *picad** gerada pelo Concord

A primeira coluna da tabela mostra as palavras que mais ocorreram nesse horizonte (p. ex., 5 palavras à esquerda e 5 à direita) por ordem decrescente de frequência (o nóculo usado na buscas do concordanciador sempre fará parte da lista). As colunas restantes mostram as posições ocupadas pelas palavras da esquerda (“L” = Left) e da direita (“R” = Right), L5, L4... L1 * R1, R2... R5 – o asterisco, neste caso, representa a coluna ocupada pelo nóculo, e só será útil nos casos em que se usa o caractere coringa na busca (como no nosso exemplo, “picad*”). Na posição em que houve maior número de ocorrências para aquela palavra, a frequência é exibida em vermelho.

Na figura 28, por exemplo, é possível observar que a palavra que mais co-ocorre com “picad-” no corpus de português é a preposição “de” (especialmente nas posições L2, L4 e L5); medidas também estão presentes nas primeiras posições, como “sopa” (de “colher de sopa”), “chá” (de “colher de chá”) e “xícara”; o ingrediente mais recorrente nesse horizonte é “cebola” (1.368 oc. à esquerda) – o que reitera a presença da forma no feminino singular, “picada”, na segunda posição da lista, com 3.706 oc. no total.

Esses dados podem ser visualizados de uma outra forma na tabela de padrões de colocados, *patterns* em inglês (vide Fig. 29) – nessa tabela, cada coluna mostra uma lista com as palavras mais recorrentes nas posições à direita e à esquerda do nóculo, por ordem decrescente de frequência¹⁵⁸. É importante ressaltar que a tabela deve ser lida apenas verticalmente (coluna a coluna), e não horizontalmente (linha a linha). Para que essa função seja ativada, primeiramente solicita-se que seja feita a lista de colocados, que é então minimizada. Em seguida, clica-se no terceiro link da direita para a esquerda da barra de ferramentas (*patterns*).

Na tela de resultados da figura 29 é possível observar, por exemplo, a ocorrência de várias ervas (salsinha, salsa, manjeriço, coentro, hortelã, alecrim) e temperos (cebola, alho) ocorrendo na primeira posição à esquerda de “picad-”, ao passo que à direita do item lexical há uma grande ocorrência de medidas (provavelmente quantificadores do ingrediente seguinte da lista).

N	L5	L4	L3	L2	L1	R1	R2	R3	R4	R5
1	de	de	sopa	de	salsinha	picada	colheres	de	de	de
2	colheres	colheres	de	cebola	cebola	picado	colher	sopa	sopa	do
3	colher	xicara	chá	a	alho	picadas	sal	chá	chá	alho
4	xicara	colher	e	o	salsa		xicara	xicara	preparo	e
5	g	sopa	dentes	sem	bem		a	e	pimenta	a
6	picada	col	xicara	cebolinha	verde		modo	colher	e	sal
7	e	e	g	e	e		g	colheres	do	azeite
8	o	chá	dente	as	média		e	col	alho	chá
9	a	g	folhas	pimentão	cebolinha	cebola	o	g	a	em
10	sopa	pele	acrescente	com	manjeriço	a	de	a	gosto	sopa
11	sem	manteiga	com	os	coentro		gosto	gosto	o	preparo
12	alho	alho	a	verde	nozes	abacaxi	dentes	pimenta	xicara	manteiga
13	chá	a	chocolate	cebolas	pequena	abobrinha	pedaços	dentes	modo	o
14	verde	ingredientes	junte	meio	sementes	abobrinhas	com	xic	sal	pimenta
15	picado	moça	moça	alho	salsão	abóbora	as	cebola	sem	queijo
16	moça	verde	verde	moça	hortelã		xicaras	o	em	picado
17	em	cebola	cebola	tomates	grande		azeite	pequenos	picada	salsinha
18	reino	o	sem	em	tomates	acerte	pimenta	sal	colheres	picada
19	manteiga	oliva	manteiga	ervas	bacon	acessório	tomates	dente	picado	modo
20	com	maço	sobremesa	azeitonas	amargo	acessórios	cebola	tomates	água	picados
21	azeite	em	adicione	tomate	fresco		preparo	do	colher	colheres
22	ingredientes	picado	pimentão	gosto	médias		cubos	modo	creme	cebolinha
23	oliva	as	alho	salsinha	tomate		pimentão	em	média	com
24	picadas	reino	o	ou	cebolas	acrescentar		com	picados	óleo
25	col	picada	talo	nozes	grosseirame	acrescente	os	ovos	grande	sem
26	as	picados	coloque	manteiga	presunto	açafrão	categoria	por	azeite	leite
27	poró	xic	tomates	queijo	alecrim	açúcar	dente	verde	caldo	farinha
28	cebola	com	óleo	cenoura	moça	adicionar	bem	bem	g	amassados
29	gosto	tomates	talos	poró	cenoura	adicione	nbsp	folhas	sobremesa	suco
30	picados	azeite	tomate	salsa	ervas		ml	xicaras	vinho	cebola
31	óleo	pimenta	salsinha	chocolate	frescas		para	fogo	queijo	vinho
32	os	óleo	poró	manjeriço	vermelho		kg	pimentão	verde	xicara

Figura 29: tabela de padrões de colocados para *picad** (*patterns*)

¹⁵⁸ aparentemente, a versão 3 do WST apresenta erro na tela de resultados para essa função – a palavra de busca aparece na coluna destinada às palavras da posição R1 (a coluna que seria a do nóculo, “*”, não é exibida na tabela).

3.3.4.5 Lista de *clusters*

Os *clusters* (ou agrupamentos lexicais, ou N-gramas) são seqüências fixas de palavras que co-ocorrem dentro de um horizonte de x palavras à esquerda e/ou à direita de um nóculo, com uma freqüência acima de y . Como a função usada para extrair os *clusters* não faz qualquer discriminação de quais palavras entram nessas seqüências, mas baseia-se somente na freqüência de co-ocorrência das mesmas, observada nas linhas de concordâncias geradas para um determinado nóculo, dentro do horizonte pré-determinado, pode haver muito ruído na lista de resultados (p. ex., "o pimentão e", ou mesmo "a cebola e" numa lista de *clusters* de 3 palavras para o nóculo "pimentão / pimentões").

É possível garantir que a palavra de busca esteja incluída em todos os itens da lista de clusters. Por exemplo, pode-se estimar o tamanho médio do tipo de agrupamento lexical que se está pesquisando – no caso de "pimentão / pimentões", seria interessante observar as palavras que co-ocorrem à direita do nóculo numa janela de até 2 ou 3 palavras, para identificar os tipos de pimentão; ou, 2 a 3 palavras à esquerda do nóculo, para identificar os verbos associados ao ingrediente. Para fazer esses ajustes deve-se acessar a tela de configurações (*Settings*) do programa (vide seta na Fig. 30) – o ideal é usar, nos horizontes (*Horizons*), um número abaixo do que foi indicado para o número de palavras no *cluster* (*Words in Cluster*); na freqüência, pode-se usar qualquer número acima de 2 (quanto maior o corpus, maior esse número pode ser).

Figura 30: ajuste dos parâmetros para a extração de *clusters*

Feitos os ajustes necessários¹⁵⁹, clica-se no botão apropriado (vide Fig. 27) para acessar o resultado. Como é possível observar na figura 31, essa estratégia pode diminuir o ruído na lista de clusters, eliminando as ocorrências que não contém o nódulo, mas não torna dispensável a análise e seleção manual das unidades válidas por parte do pesquisador.

N	cluster	Freq.
1	pimentão verde picado #	32
2	o pimentão e o	25
3	o pimentão e a	24
4	pimentão vermelho picado #	24
5	cebola e o pimentão	23
6	pimentão vermelho cortado em	22
7	pimentão verde # pimentão	16
8	pimentão e a cebola	15
9	xicara chá de pimentão	15
10	as tiras de pimentão	14
11	colheres sopa de pimentão	14
12	a gosto # pimentão	13
13	de pimentão verde picado	13
14	e o pimentão e	13
15	em rodelas # pimentão	13
16	pimentão verde cortado em	13
17	pimentão vermelho sem pele	13
18	a pele dos pimentões	12
19	de pimentão vermelho picado	12
20	picada # pimentão verde	12
21	os pimentões e a	11
22	pimentão vermelho em cubinhos	11
23	sopa de pimentão vermelho	11
24	cebola picada # pimentão	10
25	em tiras # pimentão	10
26	g de pimentão vermelho	10
27	picada # pimentão vermelho	10
28	pimentão vermelho # pimentão	10
29	tiras de pimentão e	10
30	de pimentão vermelho em	9
31	pimentão verde pequeno #	9
32	pimentão vermelho sem sementes	9

Figura 31: lista de *clusters* gerada pelo Concord

3.4 Lingüística de Corpus e Tradução

Os Estudos da Tradução Baseados em Corpora (*Corpus-based Translation Studies*, doravante CBTS) tornaram-se possíveis, na opinião de Laviosa (2002: 5), pela proximidade teórica entre as duas áreas que inspiraram e influenciaram sua formação: a L.C. e os Estudos Descritivos da Tradução (*Descriptive Translation Studies*) – uma das vertentes “puras” dos Estudos da Tradução propostas por Holmes (vide item 1.1 do Capítulo I). Essa vertente, que deveria se preocupar com a descrição sistemática do fenômeno empírico da tradução (seu produto, processo e função social), passou a ocupar uma posição de destaque nos Estudos da Tradução a partir do início da década

¹⁵⁹ o programa já vem com os parâmetros de 5 palavras para o horizonte, frequência mínima de 3 e 2 palavras no cluster, de modo que se pode usar a função sem fazer nenhum ajuste.

de 1990 – vide, por exemplo, os trabalhos de Puurtinen (1989) e, especialmente, Toury (1995).

O que o trabalho desses e de outros autores que passaram a trabalhar com a mesma linha teórica têm em comum com a L.C. é o fato de partirem da observação e descrição de corpora de textos completos autênticos (traduções), contrastando-os entre si, com seus respectivos originais ou com outros textos semelhantes produzidos por falantes nativos da língua de chegada. Ou seja, ainda que “corpus” não tenha o sentido específico que tem no âmbito da L.C., os estudos são eminentemente empíricos, e não baseados na especulação de suposições intuitivas e idéias pré-concebidas, como muitos dos estudos da área que vinham sendo feitos até então. Além disso, os pesquisadores se propunham a apresentar seus achados de forma descritiva, indicando comportamentos **prováveis**, e não regras prescritivas de como se deve ou não traduzir. Pareceria razoável, portanto, propor que as questões teóricas fundamentais da Tradução – os universais da tradução, as normas e os processos envolvidos na produção do texto de chegada – fossem estudadas usando-se as metodologias da L.C. (Laviosa, 2002: 17).

Autores como Gellerstam (1986), Church e Gale (1991), Gale e Church (1991) e Marinai *et al.* (1991) são alguns dos primeiros a apontar o uso de corpora eletrônicos para o estudo contrastivo de textos traduzidos e para o ensino da tradução. Mas Mona Baker (1993, 1995) é a responsável pelo uso pioneiro das metodologias da L.C. no estudo das características que tipificam o texto traduzido. A autora observou, por meio da análise de corpora comparáveis¹⁶⁰, que os textos traduzidos têm certas características em comum, independentemente da língua de seus respectivos originais, que os diferenciam de textos semelhantes produzidos por falantes nativos. Baker agrupa essas características em quatro categorias principais, que chama de “universais da tradução”, quais sejam:

- i. **simplificação:** tendência que os textos traduzidos têm de simplificarem o original, seja no nível lexical (uso de paráfrases), sintático (sentenças longas são subdivididas) ou estilístico (redução ou omissão de repetições);

¹⁶⁰ Mona Baker e os demais estudiosos da CBTS são os únicos a usarem o termo “*corpus* comparável” para se referirem a “two separate collections of texts in the same language: one *corpus* consists of original texts in the language in question and the other consists of translations in that language from a given source language or languages” [dois conjuntos de textos na mesma língua: um *corpus* consiste de textos na língua em questão e o outro consiste de traduções para essa língua provenientes de uma ou mais línguas] (Baker, 1993: 234); os demais pesquisadores em L.C. e Tradução usam a definição que apontamos no item 3.3.1.

- ii. **explicitação:** tendência que os textos traduzidos têm de explicitar informações que estão implícitas no original, por exemplo, quando o tradutor explica uma diferença cultural por meio de uma construção apositiva;
- iii. **normalização:** tendência que os textos traduzidos têm de se conformarem a padrões típicos da língua de chegada, por exemplo, adaptando frases do original às estruturas típicas dessa língua;
- iv. **estabilização:** tendência que os textos traduzidos têm de serem mais semelhantes entre si do que os textos produzidos por falantes nativos, que poderia ser verificada, por exemplo, por meio da observação da razão forma/ocorrência (*type/token ratio*), isto é, razão entre o número total de palavras e o número de palavras distintas de um texto.

A abordagem de Baker para estudar a tradução ganhou vários adeptos entre os pesquisadores e estudiosos da tradução (p. ex.: Maeve Olohan, Sara Laviosa, Scott Burnett e, no Brasil, Adriana Pagano, Célia Magalhães, Diva Camargo, entre outros), mas também recebeu críticas, como pontua Berber Sardinha (2002: 27), especialmente quanto ao uso do *type/token ratio* para demonstrar o fenômeno da *estabilização*, impraticável em línguas mais sintagmáticas (como o português ou o alemão), e quanto à desconsideração das influências específicas que as diferentes línguas dos originais exercem sobre o texto traduzido.

Outro fator que aproximou a L.C. e a Tradução foi a necessidade de se traduzir documentos para várias línguas por ocasião da criação da Comunidade Européia, o que impulsionou as pesquisas na área de tradução automática, em grande parte financiadas pela própria União Européia (Hunston, 2002: 123). A isso somou-se a demanda por traduções advinda do rápido avanço do processo de globalização e internacionalização do comércio, que estimulou a formação de uma vertente mais pragmática de estudos nessa interface entre L.C. e Tradução, centrada nos problemas práticos do dia-a-dia do tradutor técnico.

Essa vertente, "Applied Corpus-based Translation Studies" para Laviosa, (2002: cap. 5), dedica-se, entre outras coisas, à confecção de materiais de consulta e ferramentas de apoio à atividade tradutória, à pesquisa de metodologias de ensino de tradução e à compilação e ao uso de corpora especializados na prática tradutória (vide, por exemplo, os trabalhos de pesquisadores como Jennifer Pearson, Lynne Bowker,

Elena Tognini-Bonelli, Dorothy Kenny, Krista Varantola, Margarita Ulrych, Stig Johansson, Belinda Maia, Stella Tagnin, entre outros).

Para Berber Sardinha (2002: 18), um dos motivos que tem impedido uma aproximação ainda maior entre L.C. e Tradução é a falta de acesso à tecnologia, seja em termos de equipamentos e *software*, seja em termos de formação voltada para o uso dessas ferramentas – condição *sine qua non* para o uso da L.C. nas pesquisas com corpora:

(...) o pesquisador ou tradutor que deseje fazer incursões na exploração de corpora para a investigação da tradução enfrentará o problema da maior escassez de recursos para a sua área, da necessidade de aprender a utilizar software especializados, além de necessitar executar algumas tarefas comuns da Lingüística de Corpus, como a organização, formatação e exploração de corpus. (Berber Sardinha, 2002: 21).

Apesar das dificuldades, as pesquisas envolvendo L.C. e os mais variados aspectos da Tradução se desenvolveram bastante nesta década, inclusive no Brasil, em centros de pesquisa como o Departamento de Letras Modernas da USP (p. ex.: Orenha, 2001; Avance de Souza, 2003; Castanho, 2003; Pardo, 2004; Teixeira, 2004; Moraes, 2005; Gonçalves, 2006; Lamparelli, 2007, Carvalho, 2007; Martins, 2007; Ginezi, 2008; Fromm, 2008), o Programa de Lingüística Aplicada e Estudos da Linguagem (LAEL) da PUC de São Paulo (p ex.: Lopes, 2000; Serralvo Brit, 2004; Rossini, 2005; Alembert, 2008), a UFRGS (Salgado, 2006; Possamai, 2006; Finatto, Zilio e Fichtner, 2006; Finatto e Simioni, 2007); a UFSC (Coulthard, 2005; Gérber e Vasilévski, 2007), entre muitos outros – uma pesquisa feita em 05/10/2008 mostrou que estão registrados 16 grupos de pesquisa que trabalham com L.C. no Brasil junto à base de dados do CNPq, e uma busca por currículos na Plataforma Lattes usando as palavras de busca “Lingüística de Corpus” e “Tradução” retornou 272 pesquisadores.

Em 1999, Tagnin introduziu as primeiras noções do trabalho com corpora eletrônicos no extinto Curso de Especialização de Longa Duração em Tradução – Inglês da USP. Desde então, a L.C. vem ocupando espaço nos estudos lingüísticos da USP, bem como em outros centros de formação de Tradutores do país, tais como a UNASP, o Centro Universitário Ibero-Americano, a Uninove, a PUC-SP, a Faculdade Metodista de São Paulo e outros centros de outros estados brasileiros (tais como Acre, Brasília, Ceará, Minas Gerais, Rio de Janeiro e Rio Grande do Sul). Sua presença é

cada vez maior, também, na grande área de Lingüística e nos estudos da linguagem em geral.

Vejamos, a seguir, algumas das razões que tornam a Lingüística de Corpus uma metodologia privilegiada para o estudo e a *praxis* da tradução técnica e, por conseguinte, uma abordagem adequada para a compilação de um dicionário voltado para as necessidades de produtor de textos técnicos na língua de chegada do tradutor de Culinária.

3.4.1 Naturalidade e equivalência da perspectiva da L.C.

A naturalidade na tradução, especialmente na tradução técnica, como já foi discutido no Capítulo I, é um dos fatores que determina a qualidade do texto na língua de chegada e sua aceitação junto à comunidade de especialistas de uma dada área. Do que alguns estudiosos da L.C., da Tradução e até mesmo da Terminologia têm se dado conta nos últimos anos é que, para garantir essa naturalidade na tradução técnica, não basta ao tradutor conhecer a terminologia da área especializada, ou seja, conhecer e compreender seus conceitos numa e noutra língua. Há uma série de elementos textuais e extratextuais que participam da constituição dessa naturalidade e que, mesmo sendo essenciais na produção textual na língua de chegada – etapa fundamental da atividade tradutória –, praticamente não têm sido registrados naquele que é o principal material de consulta do tradutor: o dicionário técnico (García Palacios e Fuentes Morán, 2002: 122; Teixeira, 2004: 10; Gomez e Vargas, 2004: 4).

A questão da naturalidade do texto traduzido envolve também a questão da equivalência. Há casos em que um equivalente que poderia parecer adequado do ponto de vista terminológico-conceitual pode ser inadequado do ponto de vista textual, condicionado pela situação de comunicação em que se insere.

(...) muitas vezes, por se fazer uma tradução literal de um idioma estrangeiro para o português, acabam-se criando formas de expressão que não seriam empregadas por falantes nativos do português, embora elas sejam perfeitamente compreensíveis para eles. Algumas chegam a causar estranheza, outras passam despercebidas, embora, na prática, não se realizem. (Vasilévski, 2007: 55)

Um exemplo concreto: a figura 32 mostra um trecho de uma receita do livro *Um tesouro da cozinha clássica* (1998)¹⁶¹.

Figura 32: exemplo de tradução não-natural

Alguns “equivalentes” usados pelo tradutor, como é possível observar, prejudicam a naturalidade do texto na língua chegada. A UTE “chocolate semi-doce” (em vez de “chocolate meio-amargo”) e a abreviação “c.de chá” (“colher de chá” é, de fato, o equivalente correto de “teaspoon”, mas essa forma de abreviar não é comum em receitas), apesar de causarem um estranhamento, são compreensíveis. O mesmo não ocorre com “fermento de ação dupla” (um ingrediente ainda não disponível no Brasil) e o adjetivo “decadente”. Em inglês, o adjetivo “decadent”, quando associado a comidas, tem uma prosódia semântica positiva (“sensual”, “auto-indulgente”), já no português brasileiro, a prosódia é negativa e, portanto, totalmente inadequada para constar do título de uma receita. Ou seja, esse texto poderia ser facilmente reconhecido como um exemplar do gênero e tipo textuais que representa (receita), e pode mesmo ser compreensível. Mas, sua aceitabilidade e credibilidade junto ao público especializado (ou não) podem ser comprometidas pelas más escolhas feitas pelo tradutor e, em última análise, essas escolhas podem tornar a receita inexecutável – propósito comunicativo principal da grande maioria das receitas, e que deveria orientar todas as escolhas de equivalentes feitas pelo tradutor.

¹⁶¹ BUTLER, J. (1988). *Um tesouro da cozinha clássica*. Alberta, Canada: Creative Publishing Inc.

A equivalência, como já foi discutido também no Capítulo I, é um fenômeno condicionado pelo contexto de uso e por vezes extrapola a esfera lingüística (não pode, portanto, ser estabelecida no vácuo, num nível meramente conceitual, desvinculado do texto e das situações de comunicação reais que a conformam). Diante disso, o que torna possível a elaboração de dicionários bilíngües é a constatação, por parte dos pesquisadores em L.C., da associação existente entre **itens lexicais**, independentemente de seu status dentro de um esquema estruturado de conceitos, e certos **padrões textuais recorrentes** – estes, por sua vez, vinculados a contextos e situações de comunicação específicos e bem determinados.

Para Chesterman (1998, p. 31), equivalência é “equivalence in context”. Halliday (1992), nesse sentido, lembra que o que há são equivalentes em potencial, que por sua vez estão “contextualmente condicionados”, de tal forma que a escolha de um ou outro, na tradução, implicará na ativação de sentidos específicos oriundos da presença costumeira desse item num contexto maior. (Berber Sardinha, 2002: 31)

Como bem observa Berber Sardinha (2004: 41), o estudo da padronização da linguagem é uma das áreas que atrai mais a atenção de pesquisadores da L.C. (por exemplo: Sinclair, 1991; Hoey, 1993; Francis, Hunston e Manning, 1996 e 1998; Partington, 1998; Hunston e Francis, 2000; Berber Sardinha, 2000). No entanto, a maioria dos trabalhos disponíveis até o momento foca a língua geral e, em especial, a língua inglesa. Há poucos estudos desse tipo voltados para as linguagens especializadas, e menos ainda se consideramos a variante brasileira da língua portuguesa. No que se segue, farei um apanhado geral de como as metodologias da L.C. têm sido usadas para resolver problemas específicos da Tradução.

3.4.2 A Lingüística de Corpus a serviço da tradução

A escassez de corpora paralelos e comparáveis sempre foi apontada como um dos entraves para o estudo da Tradução via L.C. (Berber Sardinha, 2002: 20); mas, com o crescimento exponencial da Internet nos últimos 10 ou 15 anos, houve um aumento igualmente vertiginoso do volume de textos especializados disponíveis na rede, além de um avanço considerável das tecnologias de localização, coleta e processamento desses textos.

Os benefícios disso são menos evidentes para o estudo comparativo de traduções e seus originais, já que os corpora paralelos, ou quaisquer pares de textos originais acompanhados de sua(s) respectiva(s) tradução(ões), com exceção de alguns *websites* e manuais bi- ou multilíngües disponíveis *online*, continuam sendo de difícil acesso – ao menos no que diz respeito à variante brasileira do português. Além disso, não é tão fácil localizá-los na Internet, e são mais trabalhosos para processar, pois exigem, quase sempre, um pré-preparo manual dos textos, parágrafo a parágrafo ou sentença a sentença, para que possam ser alinhados automaticamente.

Para a pesquisa e o uso de corpora comparáveis na tradução, no entanto, esse aumento de informações na Internet foi um grande estímulo. Desde o final dos anos de 1990, pesquisadores de diversas instituições de ensino superior passaram a desenvolver propostas teórico-práticas voltadas para a coleta e análise desse tipo de corpus¹⁶² (p. ex.: Bowker, 1996, Maia, 1997; Kenny, 1998; Varantola, 1998; Pearson, 1998; Bernardini e Zanettin, 2000; Tognini-Bonelli, 2001; Bowker and Pearson, 2002; Tagnin, 2002a e 2002b; Olohan, 2004). Essas propostas focam necessidades específicas do tradutor, tais como o treinamento de aprendizes, a elaboração de materiais de apoio e a prática da tradução propriamente. Eis algumas metodologias de pesquisa em L.C. que podem ser usadas em corpora especializados para a solução de problemas pontuais da tradução técnica, e que usarei na identificação de equivalentes tradutórios quando da elaboração da amostra de dicionário:

- a. **caracterização de vocabulário especializado:** por meio da compilação da lista de palavras de um corpus (*wordlist*) e da lista de palavras-chave (*keyword list*) é possível obter uma lista das palavras mais freqüentes da área especializada representada pelo corpus, bem como das palavras cuja freqüência nesse corpus é estatisticamente significativa quando comparada à freqüência dessas mesmas palavras num corpus de língua geral. Também é possível observar o tamanho médio das frases e das palavras do corpus, a relação forma/ocorrência (*type/token ratio*) nos textos e no corpus total, etc. (na tela de resultados Statistics do programa WST, por exemplo);
- b. **levantamento de padrões na língua de partida:** para identificar UTEs de uma dada área técnica na língua de partida pode-se gerar as linhas de concordância

¹⁶² para uma descrição detalhada dos trabalhos em L.C. que abordam aspectos diversos da Tradução, vide Berber Sardinha (2002)

das palavras-chave do corpus (usando o programa Concord do WST, por exemplo), ou das palavras de conteúdo da lista de palavras (usando o programa WordList). Para identificar *clusters* de forma automática, pode-se fazer a lista de palavras ou de palavras-chave do corpus total usando *clusters* de 2 a 8 palavras, ou, por meio das funções *Cluster*, *Collocates* e *Patterns* da ferramenta *Concord*, acessar listas de palavras que co-ocorrem com a palavra de busca para identificar possíveis UTEs;

- c. levantamento de equivalentes em corpora paralelos:** por meio do uso de corpora paralelos alinhados é possível observar os equivalentes usados por tradutores em contextos semelhantes ao que se pretende traduzir usando-se um concordanciador paralelo, que exibe na tela todas as ocorrências de uma palavra ou expressão pesquisada alinhadas com sua(s) respectiva(s) tradução(ões) na(s) língua(s) de chegada. O corpus paralelo funciona como uma espécie de memória de tradução¹⁶³. Como bem ressalta Tognini-Bonelli (2001), nem sempre é possível ter acesso a um corpus paralelo para o levantamento de equivalentes *prima facie* (“candidatos” a equivalentes, possivelmente adequados ao contexto específico de tradução) – caso em que o uso de dicionários e da intuição¹⁶⁴ podem contribuir, desde que contrastados, posteriormente, com contextos semelhantes num corpus de textos naturalmente produzidos na língua de chegada.
- d. levantamento de equivalentes em corpora comparáveis:** quando o tradutor encontra equivalentes *prima facie* em corpora paralelos, usando dicionários e/ou sua própria intuição, pode verificar o uso desses equivalentes em corpora de textos da área produzidos naturalmente na língua de chegada, averiguando fatores como a grafia preferencial (p. ex.: “pimenta rosa” ou “pimenta-rosa”), a ordem das palavras (p. ex.: “gelatina em pó sem sabor” ou “gelatina sem sabor em pó”), a frequência (p. ex.: “capim-cidreira”, “capim-limão” ou “erva-cidreira”), o uso de preposições (“passe no / pelo ovo”), etc. Nos casos em que não é possível identificar um equivalente *prima facie* usando os métodos disponíveis, o corpus comparável ainda pode ser usado de duas formas:

¹⁶³ Bowker e Pearson (2002: 96), a propósito, vislumbram um futuro em que as memórias de tradução serão comercializadas, na forma de um *corpus* paralelo, por grandes empresas que trabalham com esse tipo de *software*.

¹⁶⁴ para uma discussão sobre o papel da “intuição” na pesquisa com corpora em tradução, vide Tognini-Bonelli (2001); Beber Sardinha (2002) e Altenberg e Grager (2002).

- **na solução de problemas pontuais:** conforme demonstra Tagnin (2007b), é possível encontrar equivalentes nos corpora da língua de chegada por meio da pesquisa dos colocados encontrados no contexto da UTE na língua de partida. Por exemplo: se o tradutor, ao verter uma receita para o inglês, não encontra um equivalente *prima facie* para a UTE “passe [INGREDIENTE] na farinha de trigo, no ovo (batido) e na farinha de rosca”, pode pesquisar as linhas de concordância das partes dessa UTE para as quais encontrar um equivalente *prima facie* (p. ex.: “flour” e “egg”). Na tela de resultados será possível identificar possíveis equivalentes para os outros itens lexicais desse padrão, como “passar” (*dip / coat / roll in*), “farinha de rosca” (*breadcrumbs*) e para a UTE como um todo, como mostra a figura 33:

Figura 33: linhas de concordância para a palavra "flour" seguida de "egg"

- **para levantamento de equivalentes *prima facie* do corpus total:** por meio da comparação das *wordlists* ou, preferencialmente, das listas de palavras-chave de corpora comparáveis, é possível identificar equivalentes *prima facie* de acordo com as posições ocupadas por essas palavras nas listas (espera-se que palavras mais recorrentes, típicas da área, ocorram mais ou menos na mesma posição nas duas listas); se uma palavra muito recorrente numa língua não apresenta um equivalente

prima facie na lista de palavras da outra língua, isso pode ser indício de que existe um problema de não-equivalência ou equivalência parcial para esse item lexical. Por exemplo: dentre as dez primeiras palavras-chave do corpus de Culinária inglês (vide Fig. 34), é possível observar que somente duas, “chopped” e “oz” (abreviação de *ounces*) não possuem pelo menos um equivalente *prima facie* na lista de palavras-chave do português. Seria interessante, nesse caso, examinar as linhas de concordância em que “chopped” ocorre para fazer um levantamento dos colocados desse adjetivo e, por meio deles, chegar aos equivalentes possíveis em português.

N	WORD	FREQ	LST %	FREQ	LST %	KEYNESS	P
1	MINUTES	12.528	0,80	770	0,01	31.159,8	0,000000
2	TBSP	10.547	0,67	0		30.833,6	0,000000
3	ADD	10.636	0,68	352		28.164,4	0,000000
4	MINS	9.264	0,59	8		27.066,2	0,000000
5	OIL	10.162	0,65	366		26.720,6	0,000000
6	UNTIL	12.083	0,77	1.842	0,04	25.420,3	0,000000
7	CHOPPED	7.843	0,50	12		22.744,9	0,000000
8	PAN	7.964	0,51	64		22.559,9	0,000000
9	OZ	7.758	0,50	12		22.496,4	0,000000
10	BUTTER	7.404	0,48	108		20.820,6	0,000000
11	INGREDIENTS	7.159	0,46	36		20.482,7	0,000000
12	HEAT	7.799	0,50	388		19.870,2	0,000000
13	SALT	7.121	0,45	180		19.212,2	0,000000
14	SUGAR	6.984	0,45	136		19.130,8	0,000000
15	PEPPER	6.574	0,42	52		18.925,6	0,000000
16	COOKING	6.731	0,43	128		18.455,7	0,000000
17	COOK	6.768	0,43	192		18.117,9	0,000000
18	TSP	6.112	0,39	0		17.854,8	0,000000
19	AND	66.899	4,27	115.135	2,23	17.245,6	0,000000
20	METHOD	6.765	0,43	560	0,01	16.104,0	0,000000
21	SERVES	5.827	0,37	148		15.712,5	0,000000
22	SERVE	6.237	0,40	428		15.267,9	0,000000
23	CREAM	5.377	0,34	80		14.913,4	0,000000
24	SAUCE	5.347	0,34	80		14.826,6	0,000000
25	TIME	11.475	0,73	6.328	0,12	13.709,3	0,000000
26	PREPARATION	5.247	0,34	210		13.656,4	0,000000
27	OVEN	4.684	0,30	28		13.351,8	0,000000
28	FLOUR	4.509	0,29	32		12.804,4	0,000000
29	GARLIC	4.445	0,28	16		12.777,7	0,000000
30	OLIVE	4.432	0,28	20		12.697,8	0,000000

N	WORD	FREQ	ST %	FREQ	T %	KEYNESS	P
1	SOPA	16.232	0,98	22		54.741,3	0,000000
2	SAL	14.087	0,85	93		46.692,3	0,000000
3	CHÁ	12.051	0,73	40		40.337,5	0,000000
4	COLHERES	10.481	0,63	0		35.527,0	0,000000
5	XÍCARA	10.316	0,62	13		34.772,5	0,000000
6	MINUTOS	11.379	0,69	507		34.590,1	0,000000
7	PREPARO	10.510	0,63	130		34.274,4	0,000000
8	COLHER	9.831	0,59	51		32.702,2	0,000000
9	INGREDIENTES	9.853	0,59	83		32.469,2	0,000000
10	COLOQUE	8.657	0,52	37		28.873,6	0,000000
11	FOGO	8.486	0,51	260		26.521,2	0,000000
12	MANTEIGA	7.673	0,46	27		25.650,0	0,000000
13	LEITE	8.111	0,49	579		23.463,5	0,000000
14	DEIXE	7.089	0,43	69		23.267,5	0,000000
15	ÁGUA	9.750	0,59	2.202	0,03	22.516,7	0,000000
16	DE	8.704	8,36	3.077	5,21	22.477,0	0,000000
17	FARINHA	6.452	0,39	53		21.263,3	0,000000
18	LEVE	6.454	0,39	163		20.400,8	0,000000
19	AÇÚCAR	7.064	0,43	514		20.382,5	0,000000
20	MOLHO	6.134	0,37	39		20.322,0	0,000000
21	MISTURE	5.988	0,36	4		20.219,1	0,000000
22	PANELA	5.996	0,36	23		20.018,4	0,000000
23	AZEITE	5.825	0,35	12		19.563,8	0,000000
24	RETIRE	5.782	0,35	6		19.493,6	0,000000
25	ATÉ	12.926	0,78	8.273	0,11	18.816,5	0,000000
26	JUNTE	5.576	0,34	8		18.769,7	0,000000
27	FORNO	6.140	0,37	272		18.658,2	0,000000
28	CEBOLA	5.475	0,33	22		18.266,9	0,000000
29	CREME	5.467	0,33	84		17.681,0	0,000000

Figura 34: primeiras palavras-chave dos corpora de Culinária de português e inglês

- e. **identificação de contextos definitórios e explicativos:** outra forma de encontrar equivalentes e/ou termos e expressões terminológicas nos corpora especializados é por meio da busca de padrões lexicais tipicamente usados para

introduzir esse tipo de informação em textos. Bowker e Pearson (2002: 219) citam os seguintes padrões em inglês¹⁶⁵:

- **descritores da relação genérico-específico:** *is a* (é um), *kind of* (espécie/qualidade de), *type of* (tipo de), *includes* (inclui / engloba);
- **descritores da relação parte-todo:** *has a* (possui/tem um), *contains* (contém/engloba), *consists of* (consiste de), *includes* (inclui);
- **descritores de função:** *used for/to* (usado para/em), *employed* (empregado em);
- **descritores da relação causa-efeito:** *causes* (ocasiona / causa), *produces* (produz / provoca), *produced by* (produzido por), *results from* (resulta de)
- **indicação de possível sinonímia:** *also called* (também chamado), *also known as* (também conhecido como), *sometimes referred to as* (às vezes chamado de).

O corpus de receitas culinárias, como já foi discutido anteriormente, não possui muitas definições, ainda que elas possam estar presentes em algumas partes da receita, como a introdução ou os comentários em notas de rodapé, mas observa-se que há várias indicações de sinonímia, relação parte-todo, etc. e até mesmo definições que são introduzidas na lista de ingredientes e no modo de fazer das receitas por meio do uso de parênteses – p. ex.: “nam pla (molho de peixe)”; “2 talos de alho-poró (somente a parte branca)”; “(...) com a ajuda de uma colher de cabo longo (bailarina)”;

- f. **na pesquisa de elementos de estilo, naturalidade e padronização:** os corpora comparáveis, ou um corpus especializado na língua de chegada, serão úteis para observar como os sentidos são tipicamente expressos em determinado gênero ou tipologia textual daquela área. Isso pode ser feito por meio da análise das linhas de concordância ou, em alguns casos, acessando contextos maiores na própria tela de resultados do concordanciador, ou diretamente no arquivo que contém um contexto particularmente semelhante ao contexto que está sendo traduzido – no programa WST, basta clicar o ícone correspondente na barra de ferramentas (vide seta na Fig. 35):

¹⁶⁵ para mais detalhes sobre essa metodologia, vide Pearson (1998).

Figura 35: exemplo de contexto expandido e acesso ao arquivo original via *Concord*

3.4.3 Lingüística de Corpus e dicionários técnicos

Como ressaltado no item 1.3 do Capítulo I, e em vários pontos deste capítulo, a maioria dos trabalhos publicados na última década que refletem sobre o uso de dicionários técnicos por tradutores, ou propõem modelos que contemplem as necessidades específicas desse usuário, é de autoria de pesquisadores ligados ao estudo, à prática ou ao ensino da tradução. E todos eles, até onde eu sei, possuem um ponto em comum: usam a Lingüística de Corpus para aproximar essas duas áreas de estudo, pois ela permite identificar em corpora de textos técnicos as informações de que o tradutor tanto precisa para a tarefa tradutória.

Varantola (1998) fez um estudo de como os dicionários são usados por tradutores e observou que, em geral, eles sentem-se frustrados com as obras de referência que consultam porque estas não lhes dão acesso a “informações que teriam sido vitais para o **processo de tomada de decisões**” (Varantola, 2002: 173, grifo meu). A autora conclui que isso acontece por vários motivos, entre eles:

- os dicionaristas, em geral, visam uma descrição do uso de uma palavra que não dependa de contextos específicos, ao passo que os tradutores buscam os dicionários justamente para solucionar o uso contextualizado de palavras;
- os tradutores em geral necessitam de informações relativas a trechos mais longos de texto, que vão além dos limites da palavra;
- tradutores buscam informações que tipicamente não constam de dicionários porque, muitas vezes, essas informações não estão disponíveis em nenhuma outra fonte de consulta.

Em trabalho posterior (Varantola, 2002), a autora vislumbra o futuro dos dicionários voltados para tradutores:

It is obvious that print dictionaries cannot cater to the particular needs of their individual users. In the future, intelligent, electronic dictionaries may approach this ideal, but I do not think that even they will manage to address the very specialized needs of their most demanding and suspicious users. I am referring to the idiosyncratic and sophisticated needs of language professionals who want to be able to make their own decisions on the basis of cumulative evidence they have gathered from different sources during their information retrieval operations.¹⁶⁶ (Varantola, 2002: 173)

Em estudo mais recente, Fromm (2008) aplicou um questionário a 177 tradutores profissionais que trabalham nas duas direções do par de línguas português – inglês sobre o uso de dicionários em sua prática tradutória. Com base nos dados levantados, o pesquisador constatou que, dentre os elementos considerados mais importantes na composição de um verbete (os elementos que tiveram acima de 70% de aprovação), os tradutores apontam, em primeiro lugar, os exemplos (84%), seguidos dos sinônimos (79%), “Palavras relacionadas / referências cruzadas (remissiva)” (77%), além de várias outras informações relacionadas à padronização textual e às instâncias macrotextuais e culturais, tais como: “Área de Especialidade” (81%), “Expressões, ditados, citações com essa palavra” (79%), “Verbos preposicionados” (79%), “Verbos

¹⁶⁶ É obvio que um dicionário impresso não pode dar conta das necessidades específicas de cada um de seus usuários. No futuro, dicionários eletrônicos inteligentes poderão chegar perto desse ideal, mas creio que nem mesmo eles poderão atender a todas as necessidades altamente especializadas de seus usuários mais desconfiados e exigentes. Refiro-me aqui às necessidades idiossincráticas e sofisticadas dos profissionais da linguagem, que esperam ser capazes de tomar suas próprias decisões com base nas evidências acumuladas em diferentes fontes de pesquisa durante uma de suas buscas por informação.

frasais (Phrasal Verbs)” (76%); “Níveis de linguagem (popular, gíria, regionalismos, etc.)” (73%) e “Pragmática (uso, campo de utilização da palavra)” (71%).

Quanto ao tipo de informação buscada, Fromm observa que 65% dos tradutores buscam, simultaneamente, a tradução (forma equivalente), uma definição e exemplos de uso. Apenas 2% buscam apenas a definição, 4% buscam somente o exemplo e 10% buscam apenas o equivalente; ou seja, o exemplo é a informação mais importante para o tradutor depois da forma equivalente. Curiosamente, Fromm não propôs a opção “tradução (forma equivalente) e exemplificação” – em minha experiência pessoal como tradutora, são as informações mínimas necessárias para a produção textual na língua de chegada. Isso se confirma nas respostas dadas a outra pergunta de seu questionário: o segundo formato de dicionário preferido pelos tradutores é aquele que “apresenta a tradução e exemplos na língua de chegada” (34%), ultrapassado apenas pelo dicionário que oferece todo tipo de informação possível (preferido por 41% dos tradutores); os que apresentam apenas a tradução (forma equivalente) ficaram em último lugar na preferência dos entrevistados (14%). Fromm observa, acertadamente, a esse respeito: “às vezes, o exemplo, associado à tradução, pode substituir a definição, considerada essencial pela maioria dos lexicólogos e terminólogos”.

A pesquisa mostra também que, na opinião dos tradutores, o material de consulta mais eficiente na resolução de problemas de tradução é a Internet (84%), seguido da consulta a especialistas (77%), dicionários impressos (76%), dicionários eletrônicos *online* (76%) e em CD-ROM (71%); a fonte que recebe menos atenção dos profissionais que responderam ao questionário é a consulta a corpora (49%). Para Fromm (2008: 63), esse resultado reflete o “estágio inicial no qual essa prática ainda se encontra e a dificuldade dos tradutores em achar instrumentos que transformem dados dispersos em um corpus eficiente de consulta”.

Por fim, a pesquisa revela que 75% dos tradutores usam 3 ou mais dicionários ao traduzir (30% usam mais de 4 dicionários), sendo que as obras mais consultadas são as bilíngües – especializadas (76%) e de língua geral (65%) – e as menos consultadas são as multilíngües (12%). Disso, Fromm concluir que:

(...) há a necessidade de busca em várias obras, o que diminui a velocidade de tradução. O profissional se beneficiaria bastante consultando apenas uma obra de conteúdo abrangente ou um local que reunisse vários tipos de obras. (Fromm, 2008: 60)

Com base nesses dados, o autor propõe o *VoTec*, Vocabulário Técnico Online¹⁶⁷, que usa a Lingüística de Corpus como metodologia na compilação dos dicionários técnicos abrigados pela *interface online* (atualmente, Computação e Lingüística, mas planeja-se incluir dicionários para todas as áreas contempladas nos corpora do CorTec, o Corpus Técnico Comparável do Projeto COMET¹⁶⁸).

O maior diferencial da ferramenta é a consulta modular, em que o usuário, de acordo com suas necessidades e seu grau de expertise na área, escolhe quais campos do verbete deseja visualizar na tela (vide Fig. 36). O pesquisador apresenta dois modelos prontos: Total (em que todos os campos são exibidos) e Tradutor (em que são exibidos os campos identificados em sua pesquisa como os mais importantes para o tradutor: definição, abreviação/acrônimo, categoria gramatical, gênero, número, ontologia, variações morfossintáticas, sinônimos, antônimos e exemplos). Em qualquer tipo de consulta, é possível visualizar a entrada com nome dos campos do verbete explicitados (consulta "Descritiva", de finalidade didática), ou apenas os dados, como numa entrada de dicionário padrão (consulta "Normal"). Outros diferenciais que poderiam ser apontados na ferramenta são:

- está disponível *online*, o que facilita a pesquisa, a atualização dos dados e a integração com outras fontes de busca disponíveis na rede – o próprio *site* oferece *links* para o "Corpus do NILC" (Lácio-Web), Google, Answers.com, Wikipedia e CorTec;
- fornece informações sobre frequência do termo no corpus e a posição que ocupa na lista de palavras por ordem decrescente de ocorrências;
- é de uso gratuito;
- sua estrutura permite que seja usado, futuramente, na construção de dicionários de outras áreas técnicas.

¹⁶⁷ <http://jr.icmc.sc.usp.br/~comet/dic>

¹⁶⁸ <http://www.fflch.usp.br/dlm/comet>

Figura 36: Tela de resultado do VoTec para o verbete “motherboard”

Em termos dos verbetes em si, da estrutura do banco de dados que os abriga e da metodologia de recolha e organização das entradas na obra, o *VoTec* usa a L.C. como metodologia, mas o modelo seguido é o tradicional: baseado numa estrutura de conhecimento da área e focado na definição e no termo. Isso o torna inadequado, por exemplo, para a presente pesquisa, especialmente porque o corpus de receitas culinárias que compilei praticamente não possui definições – a base da elaboração das estruturas de conhecimento. Outro ponto negativo do *VoTec*, na minha opinião, é o aspecto visual – no modo “Normal” da consulta tipo “Tradutor” (vide Fig. 36), por exemplo, a localização dos exemplos e, dentro desses, da ocorrência do verbete em si, é difícil, pois não é usado nenhum recurso gráfico para destacá-los do corpo da definição.

Um dos primeiros trabalhos a descrever o uso da L.C. para a localização semi-automática de contextos definitórios em corpora eletrônicos foi o de Jennifer Pearson (1998). A autora defende, entre outras coisas, que o trabalho em terminologia via L.C. exige a compilação de corpora até então preteridos por estudiosos da área – os corpora de textos de natureza mais didática, escritos por especialistas para serem lidos por não-especialistas. Segundo a autora, ao contrário dos textos altamente

especializados (escritos por especialistas para serem lidos por especialistas), esses possuem uma maior densidade de contextos definitórios e explicativos, essenciais para cumprir a tarefa de elaboração da definição terminológica de forma mais descritiva – o que sustenta a escolha teórico-metodológica de Fromm (2008), mas diverge da que será adotada neste trabalho.

Em artigo recente, *Intelligent terminological resources for translators? Content, structure and functionality* (no prelo), Temmerman defende a importância de se agregar informações ontológicas¹⁶⁹ às bases de dados e dicionários terminológicos voltados para o tradutor. Depois de analisar a importância dos contextos para o processo tradutivo, a autora propõe uma tipologia para classificá-los em três tipos: i) contextos lexicais (palavras que antecedem e se seguem à unidade de tradução); ii) contextos situacionais (conhecimentos culturais e psicológicos partilhados pelo emissor e pelo receptor no momento da criação do texto) e iii) contextos cognitivos (o estoque de memória acionado e incrementado no decurso do ato discursivo). Em seguida, ela compara a importância desses tipos de contexto para tradutores, terminólogos e termontógrafos (“termontographers”).

A autora apresenta, então, a proposta de dicionário inglês-francês de Dancette e Réthoré (2000), *Dictionnaire analytique de la distribution. Analytical dictionary of retailing*, em que as autoras procuram fornecer informações contextuais, em francês, sobre termos da língua inglesa na área de distribuição. As informações são fornecidas em vários paradigmas visualmente delimitados no verbete: *definição*, *especificações semânticas* (contexto cognitivo), *relações inter-nacionais* (contextos cognitivo e situacional-cultural), *informações complementares* (contextos lingüístico e situacional-cultural), *informações lingüísticas* (contexto lingüístico) e *exemplos* (contextos cognitivo e situacional-cultural), conforme pode ser visto no excerto da Figura 37. Uma descrição da macroestrutura do glossário pode ser encontrada em Dancette e L’Homme:

It [the dictionary] is structured around 350 key concepts that are fully described in an encyclopedic style, so to speak, but with a keen effort to highlight the semantic relationships linking the terms and to explain the nuances in meanings and regional differences in usage. In addition to the 350 full-fledged articles (the main body of the dictionary), the dictionary includes a lexicon of some 3500

¹⁶⁹ a autora considera ontologia de acordo com a aceção dada por Gruber (1993): “an ontology is a representation of the knowledge within a domain using frames and first order logic” [uma ontologia é uma representação do conhecimento no âmbito de um certo domínio usando *frames* (esquemas) e lógica de primeira ordem]

French and English related terms, covered in the body of the 350 articles.¹⁷⁰ (Dancette e L'Homme, 2002)

L1

LABEL
tag
étiquette

DÉFINITION
Document d'identification du produit qui lui est apposé ou y est attaché et qui en décrit les caractéristiques (nature, prix, provenance, marque, etc.).

PRÉCISIONS SÉMANTIQUES
Les informations contenues sur l'étiquette sont destinées au caissier (prix, codification) et au consommateur (ingrédients, mode d'emploi, etc.). L'étiquette peut présenter un graphisme plus ou moins élaboré.

Depuis les années 1970, l'étiquette comprend généralement un **code-barres** (⇒ BAR CODE). Le code-barres contient des informations telles que la description et le prix du produit, qui seront lues à l'aide d'un **lecteur optique** (⇒ OPTICAL READER).

On distingue l'étiquette **adhésive** (*self-sticking label*), ou **étiquette autocollante**, et l'**étiquette mobile** (*tag*) attachée au produit.

Sur l'**étiquette de prix** (*price tag*) ne figure que le montant à payer.

Aujourd'hui, certaines grandes surfaces n'indiquent plus le prix sur le produit mais plutôt sur la **gondole** (⇒ GONDOLA), avec des **étiquettes de gondole** (*shelf labels, shelf edge labels*) ou des **étiquettes électroniques de gondole** (⇒ ELECTRONIC SHELF LABEL).

RELATIONS INTERNATIONNELLES
Le terme anglais **TAG** désigne une **étiquette** que l'on peut facilement enlever, ce qui n'est pas le cas de **LABEL**.

Ne pas confondre l'anglais **LABEL** avec son homonyme **label**, qui a le sens de **marque** (⇒ BRAND), comme dans le terme **PRIVATE LABEL** (**marque de distributeur**).

Le terme anglais **quality label** (**label de qualité**), ou **seal of quality**, ne désigne pas l'étiquette

mais la conformité du produit à des caractéristiques ou à des qualités définies.

COMPLÉMENTS D'INFORMATION
L'idée de remplacer le **marquage** (*marking*) des prix par des codes-barres sur les produits vendus en magasin a suscité beaucoup de craintes chez les consommateurs. Ils y voyaient un préjudice à l'exercice de leur droit à l'information. Pour répondre à cette inquiétude certains distributeurs mettent des **lecteurs de codes-barres** (⇒ BAR-CODE READER) à la disposition de leurs clients, dans les divers rayons de leurs magasins. Par contre, il est admis que le code-barres présente des avantages par rapport à l'étiquette de prix, dont l'exactitude et la rapidité des enregistrements à la caisse.

Les producteurs ont l'obligation, en vertu de la **Loi sur la protection du consommateur** (*Consumer Protection Act*), de répertorier sur l'**étiquette** tous les ingrédients contenus dans le produit alimentaire.

INFORMATIONS LINGUISTIQUES
Marquage a été normalisé par l'OLF (1983) et l'ISO (1987).

étiqueter : to ticket
étiqueteuse : labeler, label machine

CONTEXTES
But it wasn't until 1900 that [he] put the first Polar **label** on a bottle of cool, naturally purified water taken directly from one of these springs on his property.
<<http://www.water.com/polar/index.html>> (30-3-99)
Dans ce but, la réglementation mise au point par les organismes de la CEE et par l'administration française prévoit sur chaque **étiquette** la présence d'un certain nombre de mentions obligatoires, en fonction de la catégorie du vin.
<<http://www.vin-champagne.com/etiq.htm>> (30-3-99)

LADS Voir LIMITED ASSORTMENT DISCOUNT STORE

Figura 37: verbete "label" em Dancette e Réthoré (2000)

O modelo de verbete proposto pelas autoras parece fazer jus à opinião de Dancette (1998 *apud* Temmerman, no prelo: 4) de que os dicionários atuais se inserem numa nova realidade pragmática, teórica e técnica que aponta para uma tendência de eliminar a distinção entre informação enciclopédica e conhecimento lingüístico, em especial nos dicionários e glossários eletrônicos. Para Dancette, graças às novas

¹⁷⁰ [O dicionário] está estruturado em torno de 350 conceitos-chave que são descritos pormenorizadamente num estilo enciclopédico, por assim dizer, mas com o firme propósito de explicitar as relações semânticas que ligam os termos e de explicar as nuances de sentido e as diferenças de uso regionais. Além dos 350 verbetes completos (o corpo principal do dicionário), a obra inclui um léxico de cerca de 3.500 termos relacionados em inglês e em francês que comparecem nas 350 entradas do dicionário.

tecnologias, “the future of specialised dictionaries for translators resides in the expansion in size of knowledge bases and in the making explicit of semantic relations in dictionaries”¹⁷¹ (*id. ibid.*).

Temmerman destaca as qualidades da obra do ponto de vista das informações contextuais oferecidas:

The dictionary compilers are experienced translators and teachers in the field of translation and terminology. At the basis of their terminological approach is their conviction that a translator with French as his mother tongue will benefit from an immersion in ontological information, i.e. information on how the terms to be translated are related to other terms in the same lexical domain or in the semantic network of terms in the French language. All this **information is given on the target language, French, in order to stimulate the discursive autonomy of the translator at the moment when he reformulates the text in the target language**¹⁷². (Temmerman, no prelo: 4, grifo meu)

Outras qualidades destacadas por Temmerman são: a consideração à estrutura prototípica do significado, o tratamento adequado da polissemia e da sinonímia, a atenção especial dada pelas autoras aos termos metafóricos da área e a organização “user-friendly” (amigável) das informações no verbete.

A meu ver, há outros aspectos positivos no dicionário de Dancette e Réthoré da perspectiva da tradução: i) os equivalentes na língua de chegada estão em negrito, seguidos do termo na língua de partida em itálico e entre parênteses – é fácil, portanto, localizá-los no texto corrido; ii) a definição, colocada logo abaixo do termo, é curta, direta e escrita em linguagem simples. Em subitens específicos, são fornecidas outras informações complementares: lingüísticas, semânticas e inter-nacionais; iii) o verbete reúne uma série de informações distintas (ou, na terminologia de Temmerman, diferentes tipos de contexto) que o tradutor, para reunir, teria que pesquisar uma gama variada de fontes de consulta; iv) o tradutor encontra informações sobre termos

¹⁷¹ *o futuro dos dicionários especializados para tradutores reside na expansão do tamanho das bases de conhecimento e na explicitação de relações semânticas em dicionários.*

¹⁷² *As compiladoras do dicionário são tradutoras e professoras experientes na área de tradução e terminologia. Na base de sua abordagem terminológica está sua convicção de que um tradutor que tenha o francês como língua materna se beneficiará de uma imersão em informações ontológicas, isto é, informações sobre como os termos a serem traduzidos se relacionam com outros termos do mesmo domínio lexical ou na rede semântica de termos da língua francesa. Toda essa informação é dada na língua de chegada, o francês, para estimular a autonomia discursiva do tradutor no momento em que ele for re-elaborar o texto na língua de chegada.*

correlatos que poderiam induzir a um erro de tradução (por exemplo, a homonímia “label → étiquette / label → marque”, em *relations internationnelles*).

Todos os aspectos destacados acima têm uma qualidade em comum – facilitam a consulta. E isso, para o tradutor, tem um sentido específico: economia de tempo. Mas, talvez, esse mesmo critério possa ser usado para erigir algumas críticas à obra. Será mesmo produtivo para o tradutor ter um verbete desta extensão? O fato de haver vários termos relacionados distribuídos ao longo dos 350 verbetes dedicados aos conceitos-chave não dificultaria sua localização, que só pode ser feita por índice remissivo? Será que as autoras tinham em mente o tradutor profissional, para quem as limitações de tempo são um fator primordial, quando da construção desse glossário? Se a extensão não era uma preocupação das autoras, porque não oferecer mais contextos de uso?

Problemas como esses poderiam ser solucionados com certa facilidade numa versão eletrônica do dicionário. Dancette teve essa mesma percepção – graças a seus esforços, a obra pode ser agora consultada na Internet¹⁷³ no *site* do DAD¹⁷⁴ *enLigne*.

Figura 38: verbete “label” no DAD *enLigne* (parte superior da tela)

¹⁷³ <http://olst.ling.umontreal.ca/dad/>

¹⁷⁴ Dictionnaire analytique de la distribution

Lá, a exemplo do que propôs Fromm (2008) para o VoTec¹⁷⁵, as consultas são modulares, isto é, o consulente opta por ampliar as informações básicas fornecidas pelo verbete inicial conforme seu interesse, necessidade e tempo disponível. Na figura 38, é possível ver a janela *pop-up* que surge ao se clicar o botão *relations internationnelles* (veja também o *link* para os contextos, no pé da tela). Se a barra lateral for rolada até o fim, é possível visualizar o seguinte quadro no verbete (vide seta):

Distribution analytique de la distribution
Jeanne Dancette

Analytical Dictionary of Retailing
Jeanne Dancette

Équipe

Recherche de termes
A B C D E F G H I
J K L M N O P Q R
S T U V W X Y Z
Retour aux types de recherche

Précisions sémantiques Relations internationnelles Compléments d'information
Informations linguistiques

Cliquez ici pour voir les contextes de LABEL (1)

Mot(s) relié(s) :

Relation sémantique :	Mot(s) relié(s) français :	Mot(s) relié(s) anglais :	Phrase source :
Instr - Instrument	étiqueteuse (1)	labeler, label machine	Source
Part - Partie de	code-barre (1)	BAR CODE	Source
Spec - Type de	étiquette adhésive (1), étiquette autocollante (1)	self-sticking label	Source
Spec - Type de	étiquette de gondole (1)	shelf label, shelf edge label	Source
Spec - Type de	étiquette de prix (1)	price tag	Source
Spec - Type de	étiquette électronique de gondole (1)	ELECTRONIC SHELF LABEL	Source
Spec - Type de	étiquette mobile (1)	tag	Source

Votre requête a donné : 15 résultat(s)
cliquez sur les termes pour accéder aux fiches
[LABEL \(1\)](#)
[LAYAWAY \(1\)](#)
[LEASED DEPARTMENT \(1\)](#)

Figura 39: verbete “label” no DAD enLigne (parte inferior da tela)

Uma busca do tipo *Requête libre* que realizei com a palavra “optical”, presente no verbete *label* (*la description et le prix du produit, qui seront lues à l'aide d'un lecteur optique* (→ *OPTICAL READER*)), dá acesso ao que seria, na versão impressa, o índice remissivo. Isso fica claro pela indicação do número das páginas, em vez de *links* para os verbetes, que constam logo após os equivalentes (em vista disso, acredito que a versão eletrônica ainda esteja em construção):

¹⁷⁵ Vocabulário Técnico Online (<http://jr.icmc.sc.usp.br/~comet/dic/?lang=pt#topo>)

Figura 40: busca livre de “optical” no *DAD enLigne*

O formato eletrônico do *DAD* parece solucionar a contento alguns dos problemas apontados no dicionário do ponto de vista da tradução: a consulta se tornou mais fácil, rápida e plástica, podendo se adaptar às necessidades individuais de cada usuário. No entanto, várias características que comentarei a seguir nos permitem afirmar que o enfoque da obra parece ser ainda muito centrada no conceito, em detrimento do **uso** dos termos em textos reais onde os sentidos são negociados e reformulados a todo momento. O *link* “Cliquez ici pour voir les contextes de BAR CODE (1)” no fim da tela (Fig. 38), por exemplo, conduz o usuário a apenas 2 exemplos de uso, um por língua – que, na verdade, cumprem apenas o papel de referenciar, abonar a fonte do termo:

Contextes :

The UPC and EAN systems utilize an agreed system of "bar codes" to enable the codes to be machine-"read" by laser scanners at the checkout. (McGoldrick 1990: 11)

[...] un bon nombre d'erreurs relevées provenaient de la négligence du commerçant à enregistrer le prix réellement en vigueur dans le programme de l'ordinateur de la caisse devant lire le code à barres imprimé sur l'emballage ou sur l'étiquette du produit acheté. (Le Soleil 5-12-97)

E mesmo os links “Source” da última coluna da tabela de relações semânticas (vide seta na Fig. 39) não levam a contextos de uso, mas às frases definitórias em que

esses termos ocorrem no próprio interior do verbete (*précisions sémantiques, relations internationnelles e informations linguistiques*) – ou seja, não há exemplos de uso para os cerca de 3.500 termos relacionados no interior dos 350 verbetes principais. Pode-se observar também que a obra debruça-se primariamente sobre os substantivos e, em menor escala, sobre os verbos (especialmente no subitem *informations linguistiques*) – mais um indicativo do enfoque semântico-conceptual.

Ressalto, mais uma vez, que minha intenção aqui não é negar a importância das informações semântico-conceptuais para o tradutor (especialmente para o tradutor que vai ao dicionário técnico em busca de uma definição, para conhecer melhor a área especializada e/ou para compreender alguma unidade de tradução do original) – aliás, as autoras trataram do tema de forma aprofundada e abrangente, apresentando as complexas relações lógico-conceptuais entre termos de forma inovadora e acessível. Mas, cumpre-nos ressaltar que muitas vezes o tradutor está em busca de outro tipo de informação.

Não há, por exemplo, no verbete de “label” quaisquer referências sobre os tipos de textos em que os termos mencionados são usados – seriam freqüentes em leis, contratos comerciais, sites institucionais, manuais de treinamento de empresas de distribuição? Certamente, cada um desses contextos de comunicação especializada possui peculiaridades que, acredito, é fundamental salientar num dicionário voltado para o tradutor. Nos casos de sinonímia, também, não há indicações de quais termos são os mais recorrentes - *etiquette adhésive / autocollante*, são intercambiáveis em qualquer situação comunicativa, em qualquer tipologia textual?

Diante do exposto, pode-se concluir que o modelo inicial de dicionário de Dancette e Réthore (2000), bem como sua versão *online*, ainda que forneçam informações muito úteis ao tradutor, especialmente ao tradutor aprendiz, talvez não seja tão eficiente na atividade tradutória, por não apresentar elementos necessários à produção textual, tais como informações lingüísticas e culturais que relacionem os termos a gêneros e tipologias textuais, informações sobre o grau de equivalência entre termos nas duas línguas, preferência de uso nos casos de sinonímia e exemplos de uso real, que permitam contextualizá-los no texto de chegada.

Essa ênfase na importância do contexto de uso para a elaboração de materiais terminológicos voltados para tradutores que venho sublinhando é ressaltada também por outros autores:

In terminology science, a term in its narrow sense designates a standardised name for a defined entity or a concept, and in its broader sense any fixed linguistic unit with a defined meaning within a special language. (...) we argue that for translation purposes the traditional understanding of terminology and terminography should be extended to include more information about term usage, collocations, variants and related concepts. Thus, some of **the phrases we present here as “useful” might not be regarded as terms in their traditional sense**. For the sake of clarity and precision it is therefore more appropriate to speak of terminological collocations.¹⁷⁶ (Vintar, 2000: 6, grifo meu)

(...) el diccionario especializado bilingüe destinado al traductor ha de contener informaciones que pueden superar los límites de lo estrictamente necesario y suficiente: ha de hacer las veces de un diccionario monolingüe en el que aparezcan una definición y contextos de uso real del término validados que aseguren la fiabilidad de la información y, asimismo, ha de contener otras informaciones como sinónimos, variaciones significativas determinadas por el contexto, categoría gramatical, relaciones conceptuales, etc., que sirvan de ayuda al traductor en **el proceso de toma de decisiones que es la traducción**. (Gomez e Vargas, 2004, grifo meu)

O que proponho, portanto, é que a L.C. seja usada para oferecer ao tradutor informações relativas à padronização de quaisquer itens lexicais cuja ocorrência seja significativa em um corpus composto de textos da área semelhantes ao que o tradutor tipicamente traduz. Além disso, dispor essas informações em um dicionário eletrônico *online* permite o acesso a vários outros recursos que podem ser úteis ao tradutor, minimizando o tempo gasto com pesquisas na rede.

Outro diferencial desta proposta será o acesso não somente às linhas de concordância previamente selecionadas durante a compilação do verbete, mas ao corpus todo, para que o tradutor possa conformar as informações oferecidas no dicionário às suas necessidades específicas de produção textual por meio da observação de outros exemplos, talvez mais próximos do seu contexto de tradução específico.

¹⁷⁶ *Em Terminologia, um termo, em seu sentido estrito, designa um denominação normatizada para uma entidade definida ou conceito, e em seu sentido lato, qualquer unidade lingüística fixa com um sentido definido no âmbito de uma linguagem de especialidade. O foco de nosso estudo, no entanto, é a terminologia voltada para a tradução, e argumentamos que, para os propósitos da tradução, o entendimento tradicional da terminologia e da terminografia deve ser ampliado de modo a incluir mais informações sobre o uso dos termos, colocações, variantes e conceitos relacionados. Assim, algumas das frases que apresentamos aqui como “úteis” pode não ser consideradas como termos em seu sentido estrito. Em favor da clareza e da precisão, seria mais apropriado, daqui em diante, falarmos em colocações terminológicas.*

A proposta do Dicionário *Online* de Culinária voltado para tradutores será formalizada no Capítulo V. Vejamos, no último item deste capítulo, como foram constituídos os corpora usados em nesta pesquisa e que estarão à disposição do tradutor durante as consultas ao dicionário eletrônico.

3.5 Compilação do corpus comparável de Culinária

3.5.1 Planejamento do corpus

Embora haja muitas propostas de classificação do texto especializado da perspectiva da Terminologia (p. ex. Picht, 2004 [2001]; Hansen, 1988; Pearson, 2004 [1999]), todas elas, diferentemente do que proponho no presente trabalho, focam a extração de termos e definições, e não de UTEs – e isso faz muita diferença para o planejamento e compilação do corpus, como bem ressalta Pearson:

Ao contrário do que se poderia pensar, nenhum texto especializado é adequado a uma extração terminográfica. Lembremos que nosso objetivo principal é selecionar textos que contenham elementos definitórios, ou seja, textos nos quais alguns termos utilizados sejam definidos de forma implícita ou de forma explícita. (...) Parece-me que toda discussão sobre o que é um texto especializado deve considerar um elemento muito importante que é extra-textual, isto é, a relação entre autor e leitor. (Pearson, 2004 [1999]: 54-55, grifo meu)

Para a autora, os textos especializados escritos por especialistas para serem lidos por especialistas não são adequados à tarefa terminológica pois possuem “uma densidade muito alta de termos, mas provavelmente muito poucos elementos definitórios” (*op. cit.*: 55). Por essa razão, dois critérios que Pearson propõe para a seleção de textos para compor um corpus terminológico é que sejam do âmbito profissional ou educativo, e que sua função principal seja informativa, didática ou normativa (p. ex.: livros voltados para a formação profissional e acadêmica).

Ora, se o objetivo é produzir um dicionário voltado para o tradutor que enfoque a padronização dos textos de maior circulação num determinado âmbito profissional (provavelmente os mais traduzidos / vertidos), são justamente os textos mais especializados que constarão do corpus a ser usado numa pesquisa com essa. Como ressalta Pearson, esse tipo de texto é inadequado à tarefa terminológica, mas,

conforme afirmado anteriormente, o objetivo aqui não é fazer terminologia (isto é, não me dedicarei à tarefa de identificar termos, conceitos e definições), mas identificar e registrar em dicionário Unidades de Tradução Especializadas, que refletem a padronização em textos técnicos – mais especificamente, em receitas culinárias.

O que diferencia a proposta desta pesquisa da de outros autores que trabalham na interface L.C. / Terminologia, a meu ver, é que para esses pesquisadores, o foco continua sendo o termo, o conceito e a definição. As metodologias da L.C. são empregadas para auxiliar o terminólogo na tarefa de identificação e descrição desses objetos de estudo que, antes, por ser manual, era mais trabalhosa. Alguns pesquisadores continuam usando o corpus apenas como um repositório de exemplos; outros, como Pearson, propõem uma abordagem diferenciada do texto técnico (privilegiando o uso de textos didáticos, por exemplo – inicialmente desprezados por terminólogos por sua baixa densidade terminológica).

A proposta desta pesquisa situa-se mais na interface entre a Lingüística de Corpus com a Tradução (e, na medida em que se almeja registrar em dicionário técnico os achados da pesquisa, com a Terminografia). Portanto, o corpus compilado para o presente trabalho é composto apenas de receitas – o tipo textual mais representativo e de maior circulação na área técnica da Culinária (vide Capítulo II). Receitas dificilmente contêm definições – os contextos explicativos, e um ou outro contexto definitório que possam existir seriam mais adequados, talvez, a um estudo diacrônico de neologismos (como o uso da palavra “vegetais” para referir-se a “legumes e verduras”, ou de “talo de erva-cidreira”), arcaísmos (como “toucinho”), empréstimos (como “nam pla”), etc.

O pesquisador que deseja compilar um dicionário para o tradutor que sirva tanto para a compreensão da terminologia veiculada pelos textos técnicos quanto para a produção textual na língua de chegada deverá compilar corpora distintos, adequados para cada uma dessas tarefas. No primeiro caso, de acordo com Pearson (2004 [1999]: 56), os critérios a serem empregados na escolha dos textos são:

- a. os textos devem ter sido **publicados** – foram escritos para serem lidos;
- b. os textos devem ser **inteiros**, pois porções distintas dos textos podem conter terminologia distinta;
- c. o **autor deve ser competente** para escrever sobre o assunto;
- d. o **âmbito** deve ser **profissional ou educativo**;

- e. a **data de publicação deve ser considerada** levando-se em conta o ritmo de evolução da área (p. ex.: informática ≠ filosofia);
- f. a **função** do texto deve ser **informativa, didática ou normativa** – por apresentar maior densidade de contextos explicativos e definitórios.

No caso de um corpus destinado ao estudo da padronização de textos técnicos com vistas à produção de um dicionário bilíngüe voltado para o tradutor, como o compilado para a presente pesquisa, proponho que a coleta de textos seja feita respeitando, **além dos critérios “b”, “c” e “e”** de Pearson (2004 [1999]: 56), os seguintes critérios:

- **os textos devem ter sido escritos por falantes nativos** (preferivelmente) – esse quesito pode ser um pouco controverso, especialmente em algumas áreas em que o descompasso tecnológico é tão grande em relação às nações mais desenvolvidas que não há textos escritos por falantes nativos sobre aquele assunto (p. ex.: nanotecnologia, quando surgiu pela primeira vez, alguns anos atrás). Excetuando-se os casos mais extremos, é um ideal a ser perseguido, pois garante a naturalidade e a autenticidade das UTEs equivalentes que comporão o dicionário;
- **os textos devem ter sido “aprovados” pelo público-alvo** preferencial das traduções típicas da área, isto é, devem ter sido publicados por editoras ou veículos de difusão (jornais, *websites*) reconhecidos por esse público; por exemplo: como o dicionário fruto dessa pesquisa trata da padronização em receitas da culinária caseira / do dia-a-dia / semi-profissional, o corpus privilegiará um tipo de texto diferente daquele requerido para a elaboração de uma obra voltada exclusivamente para a culinária profissional (p. ex.: restaurantes, formação de profissionais do setor comercial / hoteleiro, etc.) ou industrial (p. ex.: hospitais, indústria alimentícia, etc.).

Assim, a coleta dos textos para a composição do corpus comparável de Culinária desta pesquisa levou em conta o seguinte planejamento inicial:

Crítérios	Metas
língua	bilíngüe (português brasileiro e inglês, especialmente as variantes americana e britânica)
organização interna	comparável (textos originalmente escritos em português e em inglês)
tamanho	médio-grande (cerca de 1 milhão de palavras em cada língua)
meio	textos escritos (escritos para serem lidos), preferencialmente provenientes de livros ou <i>sítes</i> confiáveis de culinária
gênero	culinária caseira / semi-profissional: pratos doces e salgados, de todas as categorias (sobremesas, salgadinhos, saladas, sanduíches, massas, carnes, etc.) – evitar receitas nomeadamente “regionais”, “internacionais” e de bebidas
autoria	a mais variada possível; coletar somente receitas, no caso de <i>sítes</i> , que se supõe serem escritas por falantes nativos e terem passado por um trabalho de revisão / edição (caso sejam enviadas por internautas, por exemplo)
tipologia textual	receitas caseiras / semi-profissionais, de preferência as que têm a lista de ingredientes separada do modo de preparo
data de publicação	se for possível identificar a data de publicação, coletar receitas escritas após a década de 1990 (após a intensificação da globalização na Culinária)
nível de codificação	se possível, colocar cabeçalho em todos os textos e identificar as partes da receita (título, comentários iniciais, lista de ingredientes, modo de fazer e categorizações) com etiquetas; etiquetagem morfossintática a ser feita posteriormente

Tabela 4: critérios iniciais usados na coleta do corpus de Culinária

3.5.2 Coleta, limpeza e organização dos textos do corpus

A coleta da maioria dos textos foi feita na Internet usando-se o *offline browser*¹⁷⁷ *WinHTTrack Website Copier 3.30*, um programa distribuído gratuitamente¹⁷⁸ que faz o *download* de *sítes* e textos da Internet automaticamente para o computador do usuário¹⁷⁹.

Após um exame do conteúdo do *site* para saber se as receitas disponíveis atendiam aos critérios de coleta previamente estabelecidos, foi feita uma coleta em massa dos textos (automática) com a referida ferramenta. Uma vez arquivados em meu

¹⁷⁷ “programa feito para baixar um site inteiro de um servidor da Internet e copiá-lo fielmente no disco rígido de outra máquina. Ele foi criado com a intenção de facilitar a navegação pelos *sítes*, permitindo que o usuário navegue sem estar conectado, ou seja, que navegue *offline* (daí o seu nome)” (Berber Sardinha, 2004: 46).

¹⁷⁸ <http://www.httrack.com/>

¹⁷⁹ para mais detalhes de como usar essa ferramenta, vide Berber Sardinha (2004:45).

computador, o conteúdo dos *sites* foi inspecionado e as seções e/ou textos que não se adequavam aos critérios de coleta foram eliminados (por exemplo, fóruns de dúvidas de internautas, dicas de substituição de ingredientes, glossários, etc.).

As receitas, originalmente no formato “.html” (Hyper Text Markup Language), foram transformadas em arquivos do tipo “somente texto” (.txt), renomeadas e, em seguida, tiveram seus códigos HTML eliminados. Todas essas tarefas foram feitas usando o *Text Converter*, um utilitário do programa WST que pode ser localizado no menu *Utilities* da barra de ferramentas da tela principal do programa (vide Fig. 41).

Figura 41: como acessar o utilitário *Text Converter* do WST

Na tela principal do utilitário (Fig. 42), é possível executar várias outras tarefas, tais como: renomear arquivos, criar subcorpora a partir de um corpus maior usando campos do cabeçalho ou etiquetas previamente colocadas, corrigir erros acidentais do corpus, mudar o nome de etiquetas já inseridas no corpus, entre outras¹⁸⁰.

¹⁸⁰ Para conhecer outras possibilidades de uso desse utilitário, vide manual do usuário que acompanha o *software WST* e também Berber Sardinha (2004: 70-72) e Teixeira (2007).

Figura 42: tela principal do utilitário *Text Converter*

Maiores detalhes sobre os procedimentos específicos usados na coleta, limpeza e organização de cada subcorpora estão descritos nos itens a seguir.

3.5.3 Alguns dados gerais sobre o corpus

3.5.3.1 Subcorpus de inglês

O subcorpus de inglês é composto de 6 subcorpora, coletados em momentos diferentes e com metodologias distintas (vide Anexo 1 para um panorama geral do corpus):

1. **espec_IO:** corpus de receitas variadas coletado em 2001 por vários alunos do curso Especialização em Tradução – Inglês para a realização de um trabalho em grupo para a disciplina “Tradução e convencionalidade”. Possui receitas de todo tipo (inclusive de outros países de língua inglesa, como a Austrália) e de diferentes registros. Ao que tudo indica, a maioria das receitas provém de um extinto grupo de notícias sobre Culinária, o *rec.food.cooking* (<http://www.cs.cmu.edu/~mjw/recipes/>), mantido por Amy Gale.

→ 39 arquivos, 69.440 *tokens*, 4.796 *types*, *type/token ratio*: 6,91. Médio-baixo grau de confiabilidade;

2. **Palc_IOB:** Corpus de receitas em Inglês Original Britânico (IOB) provenientes do *Corpus Multivarietal PALC* – coletado para uma pesquisa contrastiva desenvolvida em meu mestrado¹⁸¹. As receitas foram coletadas da Internet, uma a uma, nas quatro variantes estudadas – inglês americano e britânico, português brasileiro e europeu, totalizando 440 textos (somente as 110 receitas portuguesas não foram aproveitadas na composição deste corpus). O corpus possui uma versão em que todos os textos possuem cabeçalho (vide Fig. 19, à pg. 161) e etiquetas separando as partes da receita¹⁸² (vide Fig. 43). Foram coletadas dez receitas para cada uma das 11 categorias incluídas no planejamento inicial de coleta (vide Anexo I para mais detalhes).

→ 11 arquivos, 110 receitas, 27.974 *tokens*, 2.513 *types*, *type/token ratio*: 8,98.

Alto grau de confiabilidade;

```
<titRec> Crispy Cod Balls </titRec>
<class> 30mins plus chilling; Makes approx. 40; Hot; Party Food; Buffet; Appetizer </class>
<ingr> Ingredients:
250g/9oz cod fillets, skinned
3 spring onions, chopped
2 teasp soy sauce
1 egg white
1 teasp cornflour
14 slices white bread
oil for deep frying </ingr>
<modFaz> Instructions
1. Place the fish in a food processor and blend until finely chopped.
2. Transfer the fish to a mixing bowl, add the spring onions, Soy sauce, egg white and cornflour. Mix well.
3. Remove the crusts from the bread and cut the bread into 5mm/1/4inch cubes at the largest.
4. Using damp hands, roll teaspoons of the fish mixture into balls, toss in the bread cubes, pressing firmly on, then cover and refrigerate for at least 30 minutes.
5. Preheat the deep fat fryer to 180C/350F. Deep-fry the fish balls for a few minutes until golden brown. </modFaz>
<coment> These can be prepared 1 day ahead. Make to stage 4, cover and refrigerate until required. </coment>
```

Figura 43: exemplo de receita etiquetada

¹⁸¹ Para mais detalhes sobre a compilação e exploração desse *corpus*, vide Capítulo I de Teixeira (2004) e Tagnin e Teixeira (2004a e 2004b).

¹⁸² para mais informações sobre o uso de etiquetas personalizadas em pesquisas com corpora, vide Teixeira (2007).

3. **Palc_IOA:** corpus com as mesmas características do PALC_IOB, descrito no item 2, mas com receitas escritas na variante americana do inglês.

→ 11 arquivos, 110 receitas, 30.231 *tokens*, 2.836 *types*, *type/token ratio*: 9,38. Alto grau de confiabilidade;

4. **HelenIOB:** cópia integral do site britânico *The Great British Kitchen*, editado por Helen Watson (http://www.greatbritishkitchen.co.uk/recipe_index.htm, disponível atualmente em: <http://britannia.com/cooking/recipes/>). A coleta foi feita em 2004 usando o *offline browser* WinHTTrack. A limpeza dos códigos *html* foi feita usando o utilitário *Text converter* do WST, bem como a inserção de etiquetas separando as partes da receita (ou seja, este subcorpus também possui uma versão etiquetada, mas sem cabeçalho). As receitas estão distribuídas entre 13 categorias, com números variados de receitas em cada uma (vide Anexo I para mais detalhes).

→ 13 arquivos, 1.465 receitas, 261.852 *tokens*, 6.272 *types*, *type/token ratio*: 2,40. Alto grau de confiabilidade;

5. **BBC_recs:** Receitas da seção “Food” do site da BBC, British Broadcast Corporation (<http://www.bbc.co.uk/food/recipes/>) que, além de receitas, possui seções com outras tipologias textuais – não incluídas neste corpus. As receitas são de *chefs* famosos, muitos deles com programas televisivos e de rádio na emissora, mas parece haver também receitas de internautas – ao que tudo indica, estas passam pela edição antes de serem publicadas. A coleta dos textos foi feita em 2006 usando o *offline browser* WinHTTrack, e a limpeza e renomeação dos arquivos foi feita usando o *Text Converter* do WST. Os textos não possuem cabeçalho ou etiquetas.

→ 5.180 receitas, 1.146.241 *tokens*, 9.259 *types*, *type/token ratio*: 0,81. Alto-médio grau de confiabilidade;

6. **fishSoup:** Receitas de sopa com peixes e frutos do mar coletadas manualmente na Internet, no começo de 2006, por alunos da disciplina “Linguística de Corpus e Tradução”, que ministrei no curso de Especialização em Tradução do Unibero,

para fazer a versão de uma receita semelhante. O corpus possui uma versão com cabeçalho.

→ 181 receitas, 42.387 *tokens*, 3.227 *types*, *type/token ratio*: 7,61. Médio grau de confiabilidade;

O corpus de inglês totaliza 1.578.114 *types* e 14.619 *tokens* (*type/token ratio* de 0,93). As primeiras 50 palavras-chave do corpus, quando sua Wordlist é comparada com os corpora de língua geral *Brown* (5.169.081 *types*, 47.207 *tokens*, *type/token ratio* de 0,91) e com a parte escrita do *ANC – American National Corpus* (18.472.028 *types*, 191.783 *tokens*, *type/token ratio* de 1,04), podem ser vistas do Anexo II.

3.5.3.2 Subcorpus de português

O subcorpus de português é composto de 6 subcorpora, também coletados em momentos diferentes e com metodologias distintas (vide Anexo I para um panorama geral do corpus):

1. **CDCC:** Corpus proveniente do CD “Grandes Receitas Cláudia Cozinha”, publicado e distribuído pela editora Abril, que contém 6 edições especiais de receitas da *Revista Cláudia Cozinha* (sem data de publicação), cujas receitas são “assinadas” por Betina Orrico. Por não permitir qualquer acesso aos textos a não ser via interface própria, as receitas tiveram de ser copiadas uma a uma para arquivos do tipo “somente texto”, em 27/07/2006. Para tanto, usei o programa *NoteTab Light*¹⁸³, configurado como “use as paste board” – assim, bastava-nos acionar o comando Ctrl+C na visualização “Só texto” da receita e a mesma era colada diretamente num arquivo .txt. Cada arquivo (num total de 6) contém 100 receitas de diferentes categorias (vide Anexo I para mais detalhes)

→ 6 arquivos, 600 receitas, 131.966 *tokens*, 2.804 *types*, *type/token ratio*: 2,12. Alto grau de confiabilidade;

2. **CD_Rec:** Corpus retirado do CD “1001 Receitas Práticas”, da Editora Escala (sem data de publicação), vendido com a Revista de mesmo nome em bancas de jornal. As receitas do CD são subdivididas em 17 categorias (algumas nada

¹⁸³ disponível para *download* gratuito em: <http://www.notetab.com/ntl.php> [acesso em 24/09/2008]

típicas, como “fondues, musses e suflês” – vide Anexo I para mais detalhes). Para a compilação do corpus, primeiramente as receitas em formato *.html* foram copiadas do CD, sem dificuldade. Em seguida, foi feita a conversão em lote dos arquivos (que não possuíam fotos – apenas o título da receita e os subtítulos “Modo de fazer” e “Ingredientes” em vermelho, a lista de ingredientes em negrito e um link “Clique aqui para voltar” em azul) usando-se o conversor em lote do Word do Windows. Por fim, usando a ferramenta *Text converter* do WST, eliminei os códigos *.html* restantes e criei uma versão do corpus em que o título da receita, a lista de Ingredientes e o Modo de fazer estão etiquetados.

→ 17 arquivos, 1.015 receitas, 133.340 *tokens*, 3.809 *types*, *type/token ratio*: 2,86. Médio grau de confiabilidade;

3. **palc_POB**: Corpus de receitas do Português Original Brasileiro (POB) provenientes do Corpus Multivarietal PALC (vide item 2 do subcorpus de inglês e Anexo I para mais detalhes).

→ 11 arquivos, 110 receitas, 24.816 *tokens*, 2.060 *types*, *type/token ratio*: 8,33. Alto grau de confiabilidade;

4. **Basilico**: corpus contendo receitas do site *Basilico – a gastronomia na web* (<http://basilico.uol.com.br/>), fornecidas pela editora do site à época (2003), Cristiana Couto. Possui dois subcorpora:

- i) **Bas_cabEtq**: receitas da seção “Receitas com...”: é apresentado um ingrediente e, em seguida, é fornecida uma receita em que o mesmo é utilizado. Acrescentei, manualmente, cabeçalho e etiquetas em quase todas as receitas deste subcorpus (as 16 que ficaram faltando estão no subcorpus *Bas_limpos*;

- ii) **Bas_limpos**: receitas já limpas (sem códigos *.html*), mas sem cabeçalho e etiquetas. Em sua maioria provém: a) da seção “Menu Basilico” (nome do arquivo começa com BASMB) – menus semanais, em geral temáticos, com entrada, prato principal, acompanhamento e sobremesa; b) receitas da extinta seção “receita do dia” (nome do arquivo começa com BASRD) – receitas variadas e c) 3 arquivos com várias

receitas – algumas da seção “famosos na cozinha”; algumas da seção “À mesa com Mari” e algumas receitas de restaurantes famosos.

→ 2 arquivos, cerca de 700 receitas, 193.442 *tokens*, 6.376 *types*, *type/token ratio*: 3,20. Alto grau de confiabilidade;

5. **JPAN**: coletado, em 03/08/2006, usando o *offline browser* HTTrack do site da rádio *Jovem Pan* (<http://jovempan.uol.com.br/jpamnew/destaques/culinaria/>) e limpo com o *Text converter* do WST. Consiste de receitas dadas por chefs de restaurantes famosos de São Paulo no programa de rádio *online* (no site é possível ouvir o arquivo em *Real Player*). O período das receitas compreende nov/2003 a jul/2006. Ainda é preciso colocar o nome do chef e do restaurante de proveniência das receitas, presente em praticamente todas elas, dentro de etiquetas (ou, futuramente, transferir essa informação para o cabeçalho) para evitar o aparecimento desses dados na lista de palavras do corpus.

→ 1 arquivo, 243 receitas, 41.164 *tokens*, 3.355 *types*, *type/token ratio*: 8,15. Médio grau de confiabilidade;

6. **GRGula** – receitas de chefs famosos publicadas na Revista Gula. Coletado no início de meu mestrado (2002), o corpus não possui dados sobre a data ou a localização dos arquivos na Internet. O registro parece destoar um pouco das demais receitas, apresentando um vocabulário que poderia ser associado mais à culinária profissional, seja na lista dos ingredientes, seja no modo de fazer (além dessas receitas, somente o subcorpus Basilico possui receitas mais especializadas). No entanto, por ser uma coleção bem pequena de receitas, optei por mantê-la.

→ 1 arquivo, cerca de 45 receitas, 14.046 *tokens*, 1.626 *types*, *type/token ratio*: 11,58. Médio grau de confiabilidade;

7. **Espec_PO**: Contraparte em português brasileiro do subcorpus Espec_IO, é um corpus de receitas variadas coletado em 2001 por vários alunos do curso de Especialização em Tradução – Inglês para a realização de um trabalho em grupo para a disciplina “Tradução e convencionalidade”. Possui receitas de todo tipo (inclusive algumas regionais) e de diferentes registros, contendo inclusive receitas aparentemente não revisadas (p. ex.: há uma ocorrência de “grutamato

monosódico”). Ao que tudo indica, a maioria das receitas foi retirada do site *Arroz, Feijão e Cia*: <http://www.techs.com.br/users/freibeto/>. As receitas teriam sido publicadas por um chefe chamado Norberto (freibeto), que tem um programa na TV de Araraquara.

→ 23 arquivos, cerca de 600 receitas, 102.833 *tokens*, 3.907 *types*, *type/token ratio*: 3,80. Médio-baixo grau de confiabilidade;

8. **Terra:** Corpus coletado com o *Win HTTPRack* em 08/06/2005 da seção de Culinária do portal Terra (<http://culinaria.terra.com.br/>). Os arquivos baixados, em *.html*, totalizavam cerca de 10.000 (ou seja, eram cerca de 10 mil receitas). Mas após a limpeza (feita em abril e maio de 2007) e análise da *Wordlist*, constatei que havia muitos arquivos repetidos (alguns mais de uma vez). Eliminei as duplicatas automaticamente (identificando os arquivos cujo número de *types*, *tokens* e o valor do *type/token ratio* eram idênticos), o que reduziu o corpus quase pela metade. Os arquivos restantes estão limpos e organizados. O corpus contém receitas de alta confiabilidade (em sua maioria, indicam a fonte como sendo alguma revista culinária ou agência de informação), mas há receitas de internautas (identificáveis por “Receita enviada por...”), e algumas sem nenhuma referência de autoria. Pela regularidade na ordem e no tipo de informação que aparece nos arquivos, parece-nos que este subcorpus poderia ser etiquetado (separação das partes da receita) com certa facilidade.

→ 1 arquivo, 5.024 receitas, 939.653 *tokens*, 8.500 *types*, *type/token ratio*: 0,9. Médio grau de confiabilidade;

O corpus de português totaliza 1.581.260 *types* e 13.443 *tokens* (*type/token ratio* de 0,85). As primeiras 50 palavras-chave do corpus, quando sua *Wordlist* é comparada com os corpora de língua geral *Lácio-Ref* (7.343.752 *types*, 119.756 *tokens*, *type/token ratio* de 1,63) e *Banco de Português* (230.460.560 *types*, 607.392 *tokens*, *type/token ratio* de 0,26), podem ser vistas do Anexo III.

3.5.4 Considerações finais sobre a compilação do corpus

Devido à grande dificuldade experimentada na coleta de textos da variante americana do inglês de forma automática (usando o WinHHTrack), seja por desconhecimento das ferramentas mais avançadas desse software, seja devido aos

mecanismos de proteção contra cópia presentes na maioria dos *sítes* americanos de receitas que tentei copiar, o corpus ficou com uma quantidade muito maior de textos em inglês britânico. Ainda que tenhamos observado em trabalhos anteriores (Teixeira, 2004; Tagnin e Teixeira, 2004a e 2004b) que as diferenças entre essas duas variantes do inglês não é tão expressiva em receitas quanto a que se observa, por exemplo, entre as variantes brasileira e europeia do português, o ideal seria, futuramente, balancear melhor o número de receitas em cada uma das variantes, e até mesmo incluir receitas de outras variantes do inglês, para melhorar a representatividade do corpus.

Um fator que prejudicou uma maior uniformidade na formatação dos textos do corpus (como a colocação de cabeçalhos e etiquetas separando as partes das receitas) foi a grande dificuldade que tive de fazer a limpeza dos códigos *html* de forma automática usando o *Text converter* do WST. É possível que com o uso de programas escritos exclusivamente para esse fim (em *Python* ou *Shell*, por exemplo – vide Capítulo IV) melhores resultados fossem obtidos, mas só tive acesso a essa possibilidade numa etapa posterior à compilação do corpus. Acredito que esse acréscimo de codificação ao corpus melhoraria em muito a análise ao permitir, por exemplo, um estudo do vocabulário e da padronização em setores específicos da receita, como já foi feito em corpora bem menores em outros trabalhos (vide Teixeira 2004 e 2007).

Devido a essas mesmas dificuldades de formatação automática dos textos, foi abandonado também o projeto de construir um corpus paralelo alinhado de receitas, previsto no projeto inicial do doutorado, que auxiliaria no levantamento de equivalentes *prima facie* para as UTEs identificadas em cada língua, quando da elaboração do dicionário. O corpus paralelo conta atualmente com 129.401 palavras em português original (e respectivas traduções) e 67.776 palavras em inglês original (e respectivas traduções). Está sendo trabalhado em parceria com a Languateca (Pólo Oslo) e o NILC da USP de São Carlos para ser usado com a ferramenta DISPARA (a mesma usada no site do COMPARA), e em breve será disponibilizado para pesquisa no *site* do COMET (http://www.fflch.usp.br/dlm/comet/consulta_cortrad). Para a presente pesquisa, os textos foram consultados paralelamente usando o programa de memória de tradução Wordfast (Champollion, 2001).

Passemos, agora, à etapa de exploração desses corpora.

This document was created with Win2PDF available at <http://www.daneprairie.com>.
The unregistered version of Win2PDF is for evaluation or non-commercial use only.

Capítulo IV

Compilando Unidades de Tradução Especializadas em receitas culinárias via Lingüística de Corpus

Uma das etapas importantes do trabalho terminológico consiste em descrever o perfil morfossintático-semântico das unidades especializadas da área em estudo. À parte algumas afirmações já amplamente divulgadas sobre o comportamento peculiar da língua no âmbito dos discursos especializados, como a constatação de que a grande maioria dos termos é plurivocabular (cerca de 70%) e pertence à categoria dos nomes (Krieger & Finatto, 2004: 81), cada área tem suas peculiaridades morfológicas, sintáticas, semânticas, discursivas e culturais.

Uma metodologia de extração (semi-)automática de unidades especializadas mostra-se mais eficiente quando leva em conta, além de métodos estatísticos, informações de caráter lingüístico, conforme demonstrou Teline (2004). Em seu

trabalho de mestrado, a autora conclui que os métodos de extração de terminologia baseados em dados lingüísticos (etiquetagem morfossintática, por exemplo) têm o maior índice de cobertura¹⁸⁴ (entre 90 e 100%), enquanto o método híbrido – que utiliza tanto dados estatísticos quanto lingüísticos –, é o que apresenta os melhores valores no que diz respeito à precisão¹⁸⁵.

Há muitas áreas de especialidade, no entanto – como é o caso da Culinária – em que tal perfil ainda não foi alvo de um estudo sistemático, prejudicando assim o desempenho de ferramentas de extração desse tipo, que têm de se fiar somente em informações genéricas sobre o comportamento da língua geral, ou de algumas poucas linguagens especializadas.

Ao longo dos capítulos anteriores, vim trabalhando minha hipótese de que parecem existir unidades no texto técnico, que chamei de Unidades de Tradução Especializada, que diferem das unidades em geral compiladas por dicionários técnicos terminológicos. Elas podem ser maiores do que pareceria razoável considerar como uma “entrada” de dicionário (p. ex.: “asse até que, enfiando um palito no meio, ele saia limpo”, “bata as claras em neve até obter picos moles”); podem compreender categorias até o momento desprezadas por esse tipo de obra (p. ex.: “pimenta-do-reino moída *na hora*”, “junte os ovos *um a um*”, “assadeira *não untada*”) e podem não conter termo algum e ainda serem da maior importância para garantir a naturalidade na tradução (p. ex.: “enquanto isso”, “de vez em quando”, “aos poucos”, “a gosto”).

Levando-se em conta que um dos objetivos do presente trabalho é sugerir metodologias para a extração semi-automática de Unidades de Tradução Especializada em *corpora* técnicos da perspectiva da Lingüística de Corpus, foram realizados dois estudos práticos complementares. Num primeiro momento, o objetivo era testar a hipótese de que nem todas as Unidades de Tradução Especializadas de que o tradutor necessita para a produção textual na língua de chegada são compiladas pelos dicionários terminológicos atuais. Para tanto, foi realizado um estudo mais qualitativo, que descreve e analisa manualmente a padronização de um substantivo de alta ocorrência em uma porção do *corpus* de estudo – inglês e, em seguida, contrasta

¹⁸⁴ também conhecido por “revocação” (*recall*, em inglês), é a porcentagem de respostas corretas recuperadas pelo sistema em contraposição a todas as respostas corretas possíveis, ou seja, dentre todas as respostas que o sistema deveria ter recuperado, quantas ele recuperou de fato.

¹⁸⁵ *precision*, em inglês – é a porcentagem de respostas corretas dentre todas as respostas recuperadas pelo sistema.

os dados levantados com as informações disponíveis nos dicionários técnicos da área publicados no Brasil.

No estudo seguinte, procurei caracterizar o perfil morfossintático e semântico das Unidades de Tradução Especializadas – etapa necessária para a extração semi-automática eficiente de unidades semelhantes em *corpora* eletrônicos. Realizei um estudo qualitativo-quantitativo em que são analisadas cerca de 1.300 Unidades de Tradução Especializada em português brasileiro, provenientes de minha memória de tradução Culinária sob várias perspectivas. Em seguida, os dados obtidos são testados na extração semi-automática de padrões semelhantes no *corpus* de culinária.

No que se segue, farei um apanhado da metodologia empregada e dos resultados obtidos em cada um desses estudos para, por fim, descrever os procedimentos utilizados na extração (semi-)automática, no *corpus* de Culinária, das Unidades de Tradução Especializada que futuramente comporão a amostra de dicionário técnico voltado para o tradutor fruto desta pesquisa.

4.1 Estudo da padronização de uma palavra-chave do corpus

Conforme procurei demonstrar no item 2.4.2.1, os dicionários técnicos bilíngües e multilíngües disponíveis no Brasil na área de Culinária parecem não levar em conta as necessidades dos tradutores. Dessa forma, cabe indagar: o que, afinal deve constar de um dicionário técnico bilíngüe para que seja útil na produção textual na língua de chegada – etapa fundamental da atividade tradutória? Minha hipótese é de que **o tradutor trabalha a partir de unidades textuais** (ou Unidades de Tradução Especializadas), e não a partir de palavras isoladas ou conceitos. Por isso, necessita conhecer a convencionalidade, os padrões lingüísticos recorrentes em textos autênticos da área especializada com que trabalha. A meu ver, a melhor maneira de identificar e descrever esses padrões é à luz dos pressupostos teóricos e metodológicos da Lingüística de Corpus.

Como visto anteriormente (Capítulo 3, item 3.3.3), o estudo da padronização de uma língua ou variedade lingüística visa identificar algum tipo de regularidade (lexical, sintática, semântica, etc.) na co-ocorrência de palavras. Para descrever esses padrões são usados, na literatura, três conceitos principais: i) **colocação**: associação entre itens lexicais; ii) **coligação**: associação entre itens lexicais e gramaticais; iii) **prosódia semântica**: associação entre itens lexicais e conotação (Berber Sardinha, 2004: 40).

Realizei este estudo piloto, que analisa as ocorrências do radical “-flour-” (“-farinh-”, em português), com o intuito de estudar a padronização de itens lexicais em textos culinários. Escolhi este item lexical pois, além de ser palavra-chave¹⁸⁶ da área, é um ingrediente presente na culinária mundial, nas suas mais variadas formas e composições, tanto em preparações doces quanto salgadas, seja como ingrediente principal, seja como acessório. A língua inglesa foi selecionada como ponto de partida neste estudo pelo fato de a demanda por trabalhos de tradução em Culinária ser muito maior na direção inglês-português do que na direção português-inglês, e por concordar com Baker (1992) e outros teóricos da tradução que afirmam que, em condições ideais, um tradutor deve trabalhar apenas na direção língua estrangeira → língua materna, por ser esta última a língua que, de fato, domina na condição de falante nativo.

Primeiramente, fiz um levantamento dos padrões relacionados a esse radical “-flour-” num *corpus* composto de 2.192 receitas originalmente escritas em língua inglesa – o conteúdo integral do site britânico *The Great British Kitchen*¹⁸⁷ (vide item 3.5.3.1 para maiores detalhes sobre a compilação e o composição desse subcorpus). Em seguida, procurei equivalentes para os itens levantados nos principais dicionários da área disponíveis atualmente no mercado brasileiro. O uso dos equivalentes sugeridos por esses dicionários (quando encontrados) foi atestado, então, em receitas originalmente escritas em português brasileiro usando a *World Wide Web* e o *Google*¹⁸⁸ para confirmar e complementar os achados.

4.1.1 O ponto de partida

Para identificar os padrões do radical “-flour-” presentes no corpus, primeiro passo deste estudo, foi usada a ferramenta *Concord* do programa de análise linguística *WordSmith Tools* (Scott, 1997), versão 3 (vide Fig. 44).

Digitou-se a expressão *flour* na janela de busca do programa *Concord* – os asteriscos fazem com que o programa busque todas as ocorrências em que a sequência de letras “flour” ocorre, em qualquer parte da palavra (por exemplo: **cornflour**, **flour**, **floured**, **floury**, etc.).

¹⁸⁶ Em L.C. palavras-chave são aquelas cuja frequência no *corpus* de estudo é estatisticamente significativa se comparada à frequência da mesma palavra num “*corpus* de referência” (em geral, um *corpus* de língua geral pelo menos 5 vezes maior que o *corpus* de estudo). Em outras palavras, sua ocorrência no *corpus* de estudo é muito maior do que era de se esperar caso fosse por acaso.

¹⁸⁷ <http://www.greatbritishkitchen.co.uk/cookbook.htm>

¹⁸⁸ <http://www.google.com>

Figura 44: linhas de concordância para a palavra de busca “*flour*”

Os resultados obtidos foram reordenados de várias maneiras de modo a facilitar a visualização das recorrências e co-ocorrências de palavras próximas à palavra de busca (vide seta na Fig. 44 indicando o botão que aciona essa função do programa). Em língua inglesa, os adjetivos costumam preceder os substantivos que qualificam. Assim, para visualizar os tipos de farinha encontrados no *corpus* as linhas de concordância foram reordenadas por ordem alfabética usando as palavras à esquerda da palavra de busca, procedimento aplicado ao resultado exibido na Figura 44.

Passou-se então à análise de todas as ocorrências do radical no *corpus* (2.097 ao todo) para, em seguida, tomar nota dos padrões recorrentes. Os resultados dessa pesquisa podem ser vistos nas tabelas descritas no item 4.1.2 abaixo. Na seqüência, utilizando o motor de busca *Google*, foi contabilizada a ocorrência desses padrões (quando possível) na Internet¹⁸⁹, com o intuito de verificar se não se tratava de idiossincrasias dos autores / editores do *site* escolhido para compor o *corpus*. Os

¹⁸⁹ as ocorrências oferecidas pelo Google referem-se ao número de páginas em que a palavra ou expressão de busca ocorre (de acordo com o que está armazenado em seus bancos de dados), e não o número total de ocorrências em todos os textos da Internet. Para mais informações sobre como o Google coleta e organiza essas informações, vide <http://computer.howstuffworks.com/search-engine.htm> [acesso em 06/9/2008].

resultados obtidos, bem como as expressões de busca utilizadas, estão descritos nessas mesmas tabelas.

4.1.2 Padrões de “flour” identificados no subcorpus de inglês

4.1.2.1 Padrões nominais

Na tabela a seguir, estão descritos os 29 padrões nominais (candidatos a colocação) encontrados no *corpus* de receitas culinárias originalmente escritas em inglês britânico. Alguns ocorreram apenas uma vez no corpus, mas optei por verificar a frequência de todos eles na Internet para certificar-me de que não eram limitações devidas ao tamanho reduzido da amostra ou, no caso de padrões com mais de uma ocorrência, se não eram idiosincrasias do autor / editor das receitas. Todas as frequências no *Google* exibidas nesta tabela referem-se à pesquisa feita em janeiro de 2006.

padrões nominais	no. oc. corpus	expressão de busca usada no Google	no. oc. WWW
barley flour	5	“barley flour”	40.900
cornflour ¹⁹⁰	181	“cornflour” site:.uk ¹⁹¹	18.200
floury potatoes	9	“floury potatoes” “floury potato”	834 14.300
gram flour	3	“gram flour”	18.300
granary flour	5	“granary flour”	366
malted brown flour	3	“malted brown flour”	24
malthouse flour	1	“malthouse flour”	910
plain flour	518	“plain flour” -strong ¹⁹² -unbleached	95.700
plain wholemeal flour	3	“plain wholemeal flour”	824
potato flour	1	“potato flour”	25.700
rice flour	7	“rice flour”	198.000
seasoned flour	35	“seasoned flour”	38.600
self raising ¹⁹³ flour	105	“self raising flour” -wholemeal	45.400
self-raising flour	25		

¹⁹⁰ pesquisei essa palavra por causa do homônimo, em inglês americano, “corn flour” = “fubá”

¹⁹¹ “cornflour”, em inglês britânico é o equivalente de “amido de milho” (ou “maizena/maisena”), ao passo que, em inglês americano, o mesmo ingrediente é conhecido como “cornstarch” – por isso usei o delimitador “site:.uk” para restringir a busca aos sites britânicos

¹⁹² o sinal de menos, “-”, quando colocado imediatamente antes de uma palavra de busca no Google, faz com que os resultados em que tal palavra figure sejam ignorados

¹⁹³ O Google ignora quaisquer sinais diacríticos e de pontuação, o que torna impossível a pesquisa em separado das formas “self raising” e “self-raising”. Note-se, por outro lado, que a grafia (os dois itens separados por espaço) é de uso exclusivo na culinária britânica.

self raising wholemeal flour	1	"self raising wholemeal flour"	450
soft grain flour	2	"soft grain flour"	18
sponge cake flour	1	"sponge cake flour"	25
strong plain flour	18	"strong plain flour"	952
strong white bread flour	2	"strong white bread flour"	3.170
strong white flour	7	"strong white flour"	9.420
strong wholemeal flour	4	"strong wholemeal flour"	272
unbleached plain flour	2	"unbleached plain flour"	820
unbleached plain white flour	1	"unbleached plain white flour"	12
unbleached white bread flour	7	"unbleached white bread flour"	537
wheat flour	1	"wheat flour" "wheat flour" salt OR sugar	567.000 249.000
wheatmeal flour	1	"wheatmeal flour"	178
wholemeal bread flour	3	"wholemeal bread flour"	2.380
wholemeal flour	34	"wholemeal flour" -plain -strong -self	8.430
wholemeal self raising flour	12	"wholemeal self raising flour"	1.240
wholewheat flour	1	"wholewheat flour" "whole wheat flour" ¹⁹⁴	5.620 180.000

Tabela 5: padrões nominais de "flour*"

4.1.2.2 Padrões verbais

Na tabela a seguir, estão descritos os 53 padrões verbais (candidatos a colocação e/ou coligação) encontrados no *corpus* em que o radical "flour" ocorre. Algumas convenções foram empregadas para a notação desses padrões, quais sejam:

() – elementos de uso opcional

[] – elementos de uso obrigatório

/ – elementos de uso alternado

MAIÚSCULAS – elemento que pode ser representado por diversos vocábulos da categoria / campo semântico indicado em maiúsculas

palavra ↔ palavra – posição variável da palavra (pode aparecer em qualquer uma das duas posições)

palavra ●.....● palavra – uso alternado das palavras (aparece, obrigatoriamente, numa ou na outra posição)

¹⁹⁴ O Google possui um mecanismo que permite sugerir ao usuário uma forma semelhante à pesquisada mas que possua um número de ocorrências muito maior (foi originalmente criado para evitar que os resultados obtidos a partir de expressões que, inadvertidamente, tenham erros de digitação sejam interpretadas erradamente). Quando a expressão "wholeweat flour" foi digitada, o Google sugeriu a expressão "whole wheat flour" – daí os dois resultados apresentados na tabela.

Observação: nos casos em que há mais de uma opção de vocábulo, essas aparecem pela ordem do mais para o menos freqüente no corpus.

padrões verbais	no. oc. corpus	expressão de busca usada no Google	no. oc. WWW
add (INGREDIENT(S)) to the centre of the flour	6	add "the centre of the flour" - "a well" ¹⁹⁵	49
add the flour	51	"add the flour"	92.900
blend (the) (remaining) flour [with (QUANTIFIER) / (and /,)] OTHER INGREDIENT(S)	6	"blend the flour" "blend the remaining flour"	4.670 60
coat (MEAT / FISH) in (seasoned / the) flour	10	coat +in flour ¹⁹⁶	454.000
coat (MEAT/FISH) with (a little) flour	5	coat +with flour	436.000
combine (with) (the) (remaining) flour	4	"combine flour" "combine with flour" "combine with the flour" "combine with remaining flour"	179.000 978 459 255
cut (the butter) into the flour	3	cut butter "into the flour" "cut the butter into the flour" "cut into the flour"	24.700 778 728
dip (PRONOUN / FOOD) (first) [in / into] flour, (then) (beaten) egg and (finally) (in / the) breadcrumbs	6	dip "in flour" egg breadcrumbs dip "into flour" egg breadcrumbs	6.530 1.600
dredge (PRONOUN / FOOD) (lightly) with flour	4	dredge +with flour	85.800
dust (PRONOUN / FOOD) (lightly) with (the) flour	18	dust +with flour	398.000
FLOUR – for coating / to coat	7 / 2	"flour for coating" "flour to coat" ¹⁹⁷	837 6.940
(lightly / generously) flour a [BAKING RECIPIENT / SURFACE]	4	"flour a" pan "flour a" surface	46.800 20.500
(lightly / well) floured (working) surface	162	floured surface	169.000
(lightly) floured board	28	"floured board" "lightly floured board"	77.900 29.400
[greased / buttered] [and /,] floured BAKING RECIPIENT	7	greased floured buttered floured	141.000 36.500
flour your hands	2	"flour your hands"	3.570
FLOUR, plus extra [for dusting / to dust]	3	flour "plus extra for dusting" flour "plus extra to dust"	1.100 71
FLOUR, sieved	10	"flour, sieved"	868

¹⁹⁵ usei essa expressão de busca (o sinal de menos faz com que a expressão / palavra seja desconsiderada nas buscas) porque a usada anteriormente [add "the centre of the flour"] retornou muitas ocorrências de "make a well in the centre of the flour", que é outra expressão por nós pesquisada.

¹⁹⁶ o Google despreza nas buscas, entre outras palavras de grande ocorrência, as preposições – para incluí-las obrigatoriamente na pesquisa pode-se digitar o sinal de mais (+) imediatamente antes da preposição.

¹⁹⁷ apesar do grande número de ocorrências da fraseologia "flour to coat" encontradas pelo Google, uma rápida análise das primeiras 200 ocorrências deixa ver que se trata de contexto diferente daquele em que ocorre no *corpus* – por exemplo, "Use rice **flour to coat** vegetables"; "roll in **flour to coat**"; "turning them in the **flour to coat** well".

FLOUR, sifted	14	"flour, sifted"	47.900
(gently / lightly / gradually) fold in (the) (remaining / sifted/ rest of the) flour ([and /,] OTHER INGREDIENT(S))	37	"fold in the" flour "fold in the remaining flour" "fold in the sifted flour"	66.700 475 852
knead (lightly / well / gently / quickly) on a (lightly) [floured (working) surface / board]	19	knead +on lightly floured "knead on a floured" "knead on a lightly floured" "knead gently on a lightly floured"	108.000 746 651 259
make (the) ROUX / WHITE SAUCE by (...) stirring in the flour	3	make sauce +by "stirring in the flour"	343
make [a well / a little hole] in the [centre / middle] of the flour	7	"make a well in" centre +of flour ¹⁹⁸ "make a well in" middle +of flour "make a hole in" middle +of flour "make a hole in" centre +of flour	12.600 11.300 1.890 1.690
(gradually) mix in (the) flour (and OTHER INGREDIENT(S))	8	"mix in flour" "mix in the flour"	12.200 7.460
mix (the) FLOUR [and /,] OTHER INGREDIENT(S)	18	"mix flour" "mix the flour"	90.100 33.200
mix (together) the (remaining) flour [and /,] [OTHER INGREDIENTS]	49	"mix the flour" "mix together the flour" "mix together the remaining flour"	33.200 12.000 36
mix in the flour	5	"mix in the flour"	7.460
mix together the flour, OTHER INGREDIENT(S) together (in a RECIPIENT)	28	"mix together flour" "mix together the flour" "mix the flour" together	19.100 12.000 18.800
place / put (the) (remaining) flour ([, / and] OTHER INGREDIENT(S)) in a RECIPIENT	38 / 28	"place flour" +in +a "place the flour" +in +a "put flour" +in +a "put the flour" +in +a	18.700 13.900 11.800 12.400
pour (LIQUID INGREDIENT(S)) into the centre of the flour	3	pour "the center of the flour" -"a well"	338
remaining flour	36	"remaining flour"	66.000
replace (half) the flour with OTHER DRY INGREDIENT	5	replace "flour with"	12.100
roll (MEAT / PRONOUN) in (the / seasoned) FLOUR	8	roll "in seasoned flour" roll "in the flour" "roll in the flour" "roll in flour"	3.790 38.500 761 8.890
roll out (pastry) on a (lightly) floured (working) surface	39	"roll out" +on floured surface "roll out" +on floured board	78.300 44.700
rub (the / remaining / AMOUNT) butter into (the) flour until [it / (the) mixture] resembles (fine/coarse) breadcrumbs	50	rub into flour until breadcrumbs	15.900
season (the) flour with DRY SEASONING / flour, seasoned with DRY SEASONING	3 / 1	"season flour with" "season the flour with" "flour seasoned with"	1.690 769 6.980

¹⁹⁸ apesar de os valores obtidos para as expressões com "middle / centre" serem semelhantes, vale lembrar que "centre" é uma grafia exclusivamente britânica, ao passo que "middle" resgata ocorrências em ambas as variantes do inglês: britânica e americana, o que torna o número obtido para "centre" bastante expressivo.

sift / sieve (MEASURE / QUANTIFIER) flour into [a / another] (mixing / large) bowl	6 / 1	sift "flour into" bowl sieve "flour into" bowl	16.800 6.120
sift / sieve (the) flour (and DRY INGREDIENT(S)) into a (mixing / SIZE) bowl	71 / 24	"sift flour" into bowl "sift the flour" into bowl "sieve flour" into bowl "sieve the flour" into bowl	48.600 35.800 1.130 4.540
sift / sieve <u>together</u> the flour, OTHER INGREDIENT(S) <u>together</u>	29 / 3	"sift the flour" together "sift together the flour" "sieve the flour" together "sieve together the flour"	30.500 30.500 3.890 379
sift in the (remaining) flour	2	"sift in the flour" "sift in the remaining flour"	813 77
sifted flour	3	"sifted flour"	61.100
sprinkle (the / remaining) flour into the RECIPIENT	3	"sprinkle the flour into the" "sprinkle remaining flour into"	330 3
sprinkle in the flour	6	"sprinkle in" +the flour	19.300
sprinkle with (the / a little) flour	7	"sprinkle with" flour	307.000
stir in (the) (QUANTIFIER) flour	113	"stir in the flour" "stir in the remaining flour" "stir in remaining flour"	36.100 569 5.150
stir into the flour	3	"stir into the flour"	623
stir together (the) flour, OTHER INGREDIENTS	3	"stir together flour" "stir together the flour"	33.500 10.800
tip in (all) the flour	5	"tip in the flour" "tip in all the flour"	104 44
toss (MEAT / FISH) in the (seasoned / remaining) flour	11	toss "in the flour" "toss in the flour"	17.400 254
turn ((the) dough) (out) [onto / on to] (a) (lightly) floured [(work) surface / board])	54	turn floured surface turn +onto floured surface turn +on +to floured surface	108.000 81.700 105.000
whisk in the flour	2	"whisk in the flour"	5.090
with (lightly) floured hands	5	+with floured hands	72.800
work / working in the flour	3	"work in the flour" "working in the flour"	1.400 360

Tabela 6: padrões verbais de "flour*"

4.1.2.3 Considerações preliminares sobre os dados levantados

É possível observar que todos os padrões identificadas no *corpus* tiveram uma frequência expressiva na *Internet*, com exceção, talvez, de "malted brown flour", "unbleached plain white flour", "soft grain flour" e "sponge cake flour"¹⁹⁹, que tiveram menos de 30 ocorrências. Isso parece indicar que é bem possível que um tradutor da área vá se deparar com esses padrões numa tradução e tenha que vertê-los para a língua de chegada. Resta saber que ajuda esse tradutor vai encontrar, por exemplo,

¹⁹⁹ a pesquisa foi repetida em 08/08/2008 e os resultados foram: "malted brown flour" (32 oc.), "soft grain flour" (37 oc.), "sponge cake flour" (22 oc.) e "unbleached plain white flour" (16 oc.) – "unbleached white flour", no entanto, ocorre 63.200 vezes, e "unbleached plain flour" 833 vezes.

nos dicionários bilíngües disponíveis atualmente no mercado editorial brasileiro para realizar essa tarefa.

4.1.3 Procurando equivalentes para os padrões em dicionários

Tendo feito o levantamento dos padrões contendo o radical “flour” no *corpus* de inglês, avalei, na sequência, o tipo de ajuda oferecida ao tradutor da área pelos principais dicionários de culinária disponíveis no mercado brasileiro. Maiores detalhes sobre as obras consultadas podem ser encontrados no item 2.4.2.1 do Capítulo II. Aqui, é feito um breve levantamento das possíveis soluções de equivalência fornecidas pelas mesmas caso o tradutor tivesse que traduzir especificamente os padrões relacionados a “flour” identificados no *corpus* de receitas britânicas.

4.1.3.1 Dicionários bi- e multilíngües

As obras bilíngües são o principal tipo de obra consultado numa situação de tradução. Há dois dicionários bilíngües de Culinária português-inglês disponíveis no mercado editorial brasileiro (Moura, 2000 e Klie, 2006), um no prelo (Teixeira e Tagnin), e pelo menos mais três multilíngües (que incluem o par de línguas português-inglês): Engler (2006), Saldanha (2007) e Carli e Klotz (2007).

O mais antigo deles é o *Dicionário de Culinária e Termos Afins*, de Roberto de Almeida Moura (2000). Reproduzo, a seguir, todos os verbetes – 19 no total – em que os radicais “flour” ou “farinh-” ocorrem nesse dicionário (o dicionário é idêntico²⁰⁰ na direção português-inglês, alterando apenas a língua da entrada). As entradas marcadas com asterisco possuem o radical “flour”/ “farinh-”, embora ele não esteja na entrada (ou seja, li o dicionário do começo ao fim para localizar manualmente todas as ocorrências de “*flour*” e “farinh*” na obra).

²⁰⁰ com exceção de “potato flour”, que origina duas entradas em português: “amido de batata” e “fécula de batata”

***Bread** – Cobrir com farinha de pão
***Bread crumbs** – Farinha de rosca
Corn Flour – Farinha de milho
***Cut into** – Incorporar gordura na farinha com faca, garfo ou misturador de massa
***Crumb topping** – Cobertura farinhenta, grossa
Flour – Farinha de trigo
Flour-bag – Saco para enfarinhar
Floury – Farinhento
***Graham cracker** – Bolacha meio doce de farinha integral
***Ground rice** – Farinha de arroz
Plain flour – Farinha de trigo comum
Potato flour – Amido de batata (fécula)
Roll in flour – Passar na farinha
***Roux, blonde** – Farinha cozida em manteiga clarificada, para molhos amarelo-claros
***Roux, brown** – Farinha torrada, cozida na manteiga, usada para molhos escuros
***Roux, white** – Farinha crua, cozida na manteiga, usada para molhos claros
Self raising flour – Farinha que já tem fermento
Soy flour – Farinha de soja
Wheatflour – Farinha de trigo

É possível observar várias inconsistências nos verbetes transcritos, seja em sua macro ou microestrutura. Definições inteiras constam como entrada, especialmente na direção português-inglês (**Incorporar gordura na farinha com faca, garfo ou misturador de massa** – Cut into shortening; **Bolacha meio doce de farinha integral** – Graham cracker): um indício de que a língua de partida considerada na elaboração do dicionário foi o inglês. O autor arrola como termos palavras e expressões pouco ou nada usados na área (**Saco para enfarinhar** [0 oc. na Internet] – Flour-bag; **Self raising flour** – farinha que já tem fermento [0 oc. na Internet]) – que são mais uma evidência de que o dicionário foi elaborado levando-se em consideração apenas a compreensão dos termos da língua inglesa em português. Há também informações conflitantes (**Floury** – Farinhento / **Crumb topping** – Cobertura farinhenta, grossa: deveria ser, então “floury topping”? Ou “crumby” = “farinhento?”), informações incompletas, do ponto de vista da tradução técnica, como é o caso da equivalência “**Corn Flour** – Farinha de milho” – a maioria dos dicionários de língua geral informa que “cornflour / corn flour” refere-se a ingredientes distintos na cultura americana e britânica.

O *Glossário de Gastronomia Português–Inglês / Inglês–Português* de Virginia Klie (2006), ainda que forneça equivalentes para muitos novos termos, se comparado com o dicionário de Moura (2000), apresenta vários problemas da perspectiva da tradução culinária, especialmente nos verbetes da direção português–inglês.

Reproduzo, a seguir, todas as entradas que contém o radical “flour” encontradas nesse dicionário na direção inglês-português:

(to) *bread* > cobrir com farinha de rosca
bread crumbs, of the fine dry type > farinha de rosca
cassava flour; manioc flour > farinha de mandioca
 (to) *cut in; (to) cut into* > misturar um ingrediente (manteiga) com outro (farinha de trigo) usando duas facas (cruzando-as)
 (to) *dredge* > cobrir com farinha; enfarinhar; passar na farinha; empanar
fine bread crumbs (dry type) > farinha de rosca
flour > farinha
 (to) *flour* > cobrir com farinha; enfarinhar; passar na farinha
(whole wheat) flour; graham flour > farinha de trigo integral
graham flour; whole wheat flour > farinha integral
manioc flour; cassava flour > farinha de mandioca
masa harina > certa variedade mexicana de farinha de milho desidratada
sifter > peneira para farinha

Apesar de ter 13 itens, a lista tem repetições, como “*graham flour; whole wheat flour*”, que entra tanto na ordem alfabética pela letra “g” (*graham flour*) quanto pela letra “f” (*flour, whole wheat*). Em termos de verbos, a autora apresenta quatro; “(to) flour” e “(to) dredge”, a considerar pelos equivalentes oferecidos, parecem ser sinônimos. É interessante observar a expressão usada para definir “(to) cut in / into” (= “misturar um ingrediente (manteiga) com outro (farinha de trigo) usando duas facas (cruzando-as)”) – porque a autora não usou “to cut butter in(to) flour” como entrada no inglês, se “manteiga” e “farinha” são os dois únicos ingredientes que ela arrola na explicação em português? Seria devido a uma idéia preconcebida de que entradas de dicionários não podem ser longas, ou seja, não podem ter “cara de definição”? Mas como, então, na direção português-inglês admitir uma entrada como “*certa variedade mexicana de farinha de milho desidratada*”?

A obra *Dicionário tradutor de gastronomia em seis línguas* (Saldanha, 2007)²⁰¹, por sua vez, em que os verbetes são organizados da perspectiva onomasiológica a partir do português brasileiro, o campo semântico “farinhas” apresenta 28 verbetes, dos quais 16 contêm o radical “farinh-”, em português, e 12 contêm “flour” em pelo menos um dos equivalentes em língua inglesa:

²⁰¹ a obra ainda não havia sido publicada quando da realização desta pesquisa, por isso não consta da tabela do Anexo IV.

farinha = flour
farinha de amaranto = amaranth flour
farinha de arroz = rice flour
farinha de castanha = chestnut flour
farinha de centeio = rye flour
farinha de cevada = barley flour
farinha de grão-de-bico / Besam / farinha gram = chick pea flour; besan (US) / gram flour; besan (UK)
farinha de linhaça = linseed meal
farinha de matzoh = matzo meal
farinha de milho = cornmeal (US) / maize meal (UK)
farinha de peixe = fish flour
farinha de rosca = bread crumbs
farinha de soja = soya flour
farinha de trigo duro = durum-wheat flour
farinha de trigo integral = whole wheat flour (US) / whole meal flour
farinha de trigo-sarraceno = buckwheat flour

No entanto, se consultado pelo índice remissivo, o termo “flour” aparece na lista ordenada alfabeticamente apenas em: i) “flour” – remete ao campo semântico “farinhas”, ii) “flour, to” – remete ao campo semântico “termos culinários” e iii) “floured” – também remete ao campo semântico “termos culinários”. As demais ocorrências estão listadas pela ordem alfabética do modificador (por exemplo: “barley flour” consta na letra “b”, “chestnut flour” na letra “c”, etc.), o que dificulta enormemente a pesquisa onomasiológica na ordem reversa (direção inglês – português). O único verbo ou locução verbal presente neste dicionário com o radical “flour” é “to flour” – se há outros, não é possível recuperá-las usando os índices inicial e remissivo da obra.

4.1.3.2 Dicionários monolíngües

Os dicionários técnicos monolíngües também são uma fonte de consulta importante para o tradutor. Em trabalhos anteriores (Teixeira, 2003 e 2004), demonstrei os principais problemas das obras disponíveis à época no Brasil (Gomensoro, 1999; Algranti, 2000; Vieira e Cândido, 2000; Fornari, 2001 e Jones, 1996) na qualidade de fontes de consulta para tradutores valendo-me principalmente dos parâmetros teóricos da Terminologia, da Tradução e da Lingüística de Corpus. Aqui, procuro apontar, de uma perspectiva mais prática, se e em que circunstâncias essas obras teriam sido úteis numa situação de tradução e, mais especificamente, na tradução de padrões que contêm o radical “flour”.

Como essas obras são muito mais volumosas que as obras multilíngües, não serão discutidas em detalhe todas as ocorrências do radical “farinh-” que delas constam. Serão utilizadas apenas para confirmar o registro dos equivalentes sugeridos pelos dicionários bilíngües, quando houver.

4.1.4 Cotejando os dados

A etapa seguinte da pesquisa consistiu no cotejamento dos dados obtidos no *corpus* de inglês, já confirmados na Internet, com os dicionários bilíngües de Moura e Klie para levantamento de possíveis equivalentes para os padrões relacionados a “flour” encontrados no *corpus*. Em seguida, verifiquei, para cada equivalente levantado, se o mesmo constava dos dicionários monolíngües. Por fim, pesquisei sua ocorrência na Internet, em receitas originalmente escritas em português brasileiro, usando o motor de buscas *Google*. O Anexo IV reproduz a tabela resultante dessa pesquisa. A seguir, apresento dois exemplos ilustrativos: *potato flour* e *(lightly) floured [(work) surface / board]*.

Um exemplo: tomei primeiramente a colocação “potato flour” e consultei os dicionários bilíngües e multilíngües: em Moura (2000) o equivalente oferecido é “amido de batata (fécula)”, em Klie (2006) o termo não consta, e em Saldanha consta no índice remissivo do inglês apenas “potato starch” (que remete para o verbete de “fécula de batata”). O passo seguinte foi consultar os dicionários monolíngües para verificar se os equivalentes sugeridos por Moura e Saldanha constavam de seus verbetes, e se haveria outros sinônimos.

Em Gomensoro (1999) consta apenas o termo genérico “amido” e, em “fécula” lê-se: “amido que pode ser obtido da batata (...)”. Em Algranti (2000) consta somente o “amido de *milho*”, mas “fécula de batata” comparece como verbete; em Vieira e Cândido (2000) constam apenas os termos genéricos “amido – (...) obtido de plantas e grãos de cereais como trigo, milho, arroz, batata (...)” e “fécula”; em Fornari (2001) constam “amido – (...) encontra-se, também, nos feculentos, como a batata (...)” e o genérico “fécula”. Enfim, a maioria dos dicionários confirma a existência de ambos os equivalentes sugeridos, mas nenhum aponta para uma preferência de um ou outro (e o tradutor terá de tomar essa decisão, quando for produzir o texto na língua de chegada).

Foi feita então uma pesquisa no *Google*, com o intuito de atestar o uso dos dois equivalentes identificados em todas as obras consultadas, e o resultado foi, em janeiro

de 2006: “amido de batata” = 146 oc. e “fécula de batata” = 1.110 oc. A pesquisa foi repetida em 12/6/2008 e os resultados foram: “amido de batata” = 857 oc. e “fécula de batata” = 74.100 oc. O resultado parece indicar que a preferência por “fécula de batata” confirmou-se no decorrer desses dois anos. Pode-se concluir que, atualmente, o melhor equivalente tradutório para “potato flour” é “fécula de batata”.

Por sua vez, “fécula de batata” consta como tradução de “potato starch” no dicionário de Saldanha (2007) – a autora não menciona “potato flour”. Ao serem pesquisadas no *Google* (acesso em 12/6/2008) as frequências de “potato flour” (132.000 oc.) e “potato starch” (713.000), observa-se que o segundo parece ser o equivalente tradutório mais usado em língua inglesa para “fécula de batata” – mesmo utilizando o delimitador “site:.uk”, o resultado é três vezes maior para “starch”. O ideal, no entanto, seria repetir essas buscas em um *corpus* apenas de receitas e em que se possa diferenciar as variantes do inglês para corroborar tal informação.

Passemos agora aos padrões verbais. Como já foi mencionado, as únicas entradas contendo verbos que constam dos dicionários bilíngües disponíveis no Brasil são: **“Roll in flour – Passar na farinha”** e, indiretamente, **“Cut into – Incorporar gordura na farinha com faca, garfo ou misturador de massa”** (Moura, 2000); “(to) bread > cobrir com farinha de rosca”, “(to) cut in; (to) cut into > misturar um ingrediente (manteiga) com outro (farinha de trigo) usando duas facas (cruzando-as)”, “(to) dredge > cobrir com farinha; enfarinhar; passar na farinha; empanar” e “(to) flour > cobrir com farinha; enfarinhar; passar na farinha” (Klie, 2006); “to flour” e “floured” (Saldanha, 2007), ainda que, *stricto sensu*, não possam ser consideradas colocações nesta obra, pois não apresentam os padrões propriamente e/ou as palavras com que co-ocorrem. Como o tradutor poderia proceder para traduzir, por exemplo, um trecho como: “Turn dough out onto floured surface and knead gently until smooth and elastic” (em que se pode identificar uma ocorrência do padrão: turn ((the) dough) (out) [onto / on to] (a) **(lightly) floured [(work) surface / board]**)?

Em Moura (2000), não há verbete para “to turn” ou “to turn out (onto / on to)”, assim como não há verbete para “work surface”, “surface”, “light” (apenas *light cream* → creme de leite) ou “floured” (ainda que *flour* e *floury* constem – o tradutor, se não tivesse familiaridade com a área, deduziria *flour* → farinha / *floured* → enfarinhada?). O dicionário traz a tradução para “board” (= *tábua*) que, nesse contexto especificamente, não seria um bom equivalente tradutório. Apenas a equivalência “dough → massa” oferecida por esse dicionário seria proveitosa na tradução desse padrão.

Em Klie (2006) encontra-se “(to) *turn* > virar” mas, curiosamente, o equivalente para “(to) *turn out*” é “colocar” (uma consulta ao *corpus* de português mostra que “transfira para” e “vire sobre” seriam os equivalentes mais adequados para “*turn (out) [on to / onto]*” nesse contexto). No verbete “dough”, o equivalente oferecido é “massa crua de pão, pizza ou torta” – ainda que a explicação seja válida e útil, não é um “equivalente tradutório” *per se*, ou seja, o tradutor não poderia usar o equivalente, tal qual oferecido pelo dicionário, numa tradução (partido-se do pressuposto que um dicionário desse tipo, com uma lista *termo LP* → *termo LC*, oferece equivalentes, e não definições / explicações). O dicionário não apresenta verbetes para “light” e “working surface”, mas oferece os equivalentes “*surface* > superfície” e, indiretamente, para “floured”, já que no verbete “(to) flour” consta “enfarinhar” como equivalente.

O dicionário multilíngüe de Saldanha (2007) foi pesquisado pelo índice remissivo inglês – português e foram encontrados:

- “turn out, to”, que remete ao campo semântico *Termos culinários*, verbete 2425, “**Desenformar**: [IN] unmould, to; turn out, to” – não seria de nenhuma ajuda no caso do contexto pesquisado;
- “dough” remete ao campo semântico “Outros ingredientes”, mais especificamente ao verbete 1808, que é idêntico ao equivalente oferecido pelo dicionário de Klie (2006): “**Massa crua de pão, pizza ou torta**” – que, em si, não seriam bons equivalentes para “dough” no contexto que estamos tentando traduzir, mas serve-nos para apontar a equivalência “dough → massa”;
- “light” aponta para dois verbetes do campo semântico *Termos culinários*: 2316, “**Baixas calorias**” e 2575, “**Leve**” – o segundo oferece alguma ajuda para encontrar o equivalente “levemente”, mas, por não estar na forma de advérbio, mas sim adjetivo, não faz qualquer menção à forma concorrente, também muito usada neste contexto: “ligeiramente”;
- “floured” aponta para o verbete 2464, “**Enfarinhado**: [IN] floured; dredged” – um bom equivalente para “floured” no contexto estudado, mas não seria a única opção caso estivéssemos traduzindo “Put the circles on an oiled, floured baking sheet” em que “floured” poderia ser traduzido também pela forma concorrente e também muito usada, “polvilhad* com farinha”;
- “(working) surface” e “board” não constam deste dicionário.

É possível que o tradutor, tendo consultado essas três obras e “juntado os pedacinhos” oferecidos por cada uma delas, fosse capaz de traduzir essa unidade de tradução especializada – especialmente se já tiver tido um pouco de contato com textos da área em português que tratam da preparação de massas levedadas. Mas, para finalizar sua pesquisa a contento e confirmar o equivalente “composicional” deduzido a partir dessas obras bilíngües, muito provavelmente o tradutor irá (ou deveria) consultar a Internet, ou um *corpus* especializado. Ora, se o padrão “turn ((the) dough) (out) [onto / on to] (a) (lightly) floured [(work) surface / board]” ocorre 54 vezes apenas nesse *subcorpus* de cerca de 300.000 palavras, porque não incluí-lo como entrada num dicionário voltado para o tradutor?

4.1.5 Considerações finais sobre o estudo

Essa investigação inicial, acredito, corrobora minha hipótese da importância do estudo da padronização textual para a construção de dicionários especializados que sejam mais úteis à tarefa de tradução, especialmente no que diz respeito à produção textual. O levantamento dos 82 padrões de que o radical “flour” participa em um *corpus* de 2.192 receitas originalmente escritas em inglês britânico evidenciam a insuficiência de informações a que está sujeito o tradutor brasileiro da área quando esses dados são contrastados com os verbetes em que os radicais “flour” e “farinh-” ocorrem nos dicionários bilíngües de culinária disponíveis no mercado editorial brasileiro atualmente.

Por outro lado, o uso da Internet para verificar o número de ocorrências dos equivalentes levantados nesse e nos demais dicionários em textos de Culinária originalmente escritos em português brasileiro parece indicar que, muito embora essas obras ofereçam, sim, algum tipo de ajuda ao tradutor na compreensão do texto original, muitas vezes os equivalentes sugeridos não são de uso corrente em textos autênticos da área. Além disso, a pesquisa de estruturas maiores e com itens lexicais variados (p. ex. “roll (MEAT / PRONOUN) in (the / seasoned) FLOUR”) usando o Google mostrou-se deficitária, já que esse mecanismo de busca não foi construído para esse fim – o que aponta para a necessidade do uso de *corpora* para pesquisas lingüísticas.

Quanto à metodologia, mostrou-se morosa e bastante trabalhosa. Mas, em situação de trabalho ideal, haveria dois *corpora* comparáveis – ou seja, semelhantes em tamanho, tipologia, gênero, etc. – nas duas línguas e, somente após o levantamento dos padrões textuais em ambos (que seria feito o máximo possível de

forma automática ou semi-automática, usando as metodologias computadorizadas da L.C.), procederia-se ao estabelecimento das equivalências, minimizando, assim, a consulta aos dicionários para levantamento de equivalentes. Esses seriam utilizados, bem como a consulta a especialistas da área, para sugerir possíveis substitutos e explanações quanto aos termos sem equivalente nas culturas de partida ou de chegada, ou para dirimir dúvidas nos casos em que haja equivalência parcial entre as línguas.

Tendo em vista que seria impraticável realizar esse mesmo procedimento para pesquisar cada uma das palavras-chave do *corpus* de culinária, caso se optasse por elaborar uma obra abrangente da área, o passo seguinte da pesquisa consistiu em estudar formas de extrair padrões, como os que foram identificados manualmente para “flour”, num *corpus* especializado da área de forma automatizada ou semi-automática. Esse foi o objetivo do segundo estudo, descrito no item a seguir

4.2 Estudo das Unidades de Tradução provenientes da memória de tradução

4.2.1 O ponto de partida

Muito tem sido feito nas áreas de Terminologia, Processamento da Linguagem Natural e Lingüística Computacional para a extração automática de termos em *corpora* especializados. Mas, a hipótese central deste trabalho é a de que as Unidades de Tradução Especializadas (UTES) vão além de termos ou, melhor dizendo, vão além do que a maioria dos dicionários técnicos costuma incluir em suas entradas, conforme procurei demonstrar no estudo descrito no item 4.1, em que foi feito um levantamento manual de todos os padrões de “flour” em parte do *corpus* de inglês – foram identificados 82, em sua grande maioria ausentes nas obras bilíngües e monolíngües consultadas. A metodologia empregada, embora eficiente, mostrou-se trabalhosa e demorada. Me perguntei: como identificar padrões como esses, com os quais o tradutor se deparará constantemente em seu trabalho de tradução, de forma automática ou semi-automática em *corpora* eletrônicos, para que possam ser incluídos em um dicionário especializado voltado para o tradutor?

Para tentar responder a essa pergunta, planejei este segundo estudo, que visa caracterizar UTES da área da Culinária quanto ao seu perfil morfossintático e semântico. O ponto de partida é uma lista de UTES mono- e plurivocabulares em

português com cerca de 1.300 itens, provenientes do glossário que venho compilando, desde 2004, na memória de tradução do *WordFast*[®] (Champollion, 2001), com base nos trabalhos de versão e tradução de textos culinários feitos por mim desde então.

O glossário consiste em um arquivo do tipo somente texto (.txt), com valores separados por marcas de tabulação – termo em português, equivalente em inglês, comentários (vide Fig. 45). Não usei qualquer critério na compilação dessa lista – reuni todos os glossários de Culinária que tinha em meu computador até aquele momento (trabalhos apresentados no curso de Especialização em Tradução, algum material retirado da Internet, etc.) e apenas fui acrescentando, manualmente, as unidades de tradução que iam ocorrendo em meus textos de trabalho (quando me lembrava de fazê-lo, devo admitir). Assim, sabemos tratar-se de inventário não representativo – o que deve ser levado em conta ao serem feitas generalizações a partir dos dados levantados. Mas meu objetivo aqui, como já mencionei, é fazer uma primeira aproximação dessas UTEs para, em seguida, incorporar os resultados às metodologias de busca semi-automatizadas quando da exploração do *corpus* de estudo.

Para começar a trabalhar com os dados do glossário, foi copiado todo o conteúdo do arquivo para uma tabela do *Excel*[®]. Com isso, as informações ficaram organizadas em colunas. As duas colunas da direita (equivalente e comentários) foram então eliminadas, restando apenas a primeira, que continha as unidades de tradução em português. Em seguida, foi feita uma limpeza superficial na lista, retirando unidades que haviam sido acrescentadas ao glossário apenas para evitar a digitação durante a tradução (como: “100 g (3.5 oz)”, ou “além de X horas de freezer” – estrutura típica de um livro em particular que traduzi).

A partir daí, foram criadas várias colunas para abrigar os dados obtidos nas pesquisas que se pretendia efetuar, a saber:

- qual é a categoria gramatical de cada palavra que compõe as UTEs (unidades mono- e plurivocabulares)?
- qual é a função sintática da UTE como um todo?
- quantos elementos há na UTE?
- a UTE contém alguma palavra-chave²⁰²?
- qual a frequência da UTE no *corpus* total²⁰³?

²⁰² para maiores detalhes sobre os parâmetros empregados no levantamento das palavras-chave, vide item 4.2.3.

Figura 45: glossário proveniente da memória de tradução²⁰⁴

Tendo analisado as respostas obtidas para essas perguntas, o passo seguinte foi efetuar a extração automática, numa versão etiquetada do *corpus* de Culinária em português, de todas as UTEs correspondentes aos padrões morfossintáticos identificados. Os dados resultantes foram abrigados numa nova tabela Excel e quantificados. A precisão do método de extração foi analisada em amostras aleatórias e, em vista dos resultados obtidos, foram feitas melhorias nos mecanismos de busca automática. Os resultados da extração servirão de base para a elaboração da amostra de glossário, que será abordada no Capítulo V.

4.2.2 Associação entre categorias gramaticais

Partindo da hipótese de que existe um padrão de associação entre categorias gramaticais e itens lexicais que tipifica as UTEs presentes em receitas culinárias, o primeiro passo dado para responder às perguntas iniciais deste estudo foi classificar as UTEs provenientes de minha memória de tradução quanto à categoria gramatical de cada um de seus elementos e do padrão como um todo. Em vista do grande volume de dados, optou-se por fazer essa classificação das unidades usando um etiquetador automático.

²⁰³ maiores detalhes sobre a composição do *corpus* estão descritos no item 3.5 do Capítulo III.

²⁰⁴ os zeros na figura indicam que não há comentários para aquela entrada do glossário.

4.2.2.1 Procedimento de etiquetagem

Para fazer a etiquetagem automática das unidades em estudo, ou seja, atribuir uma categoria gramatical a cada palavra da lista de forma automática usando um programa computadorizado, foi utilizado o etiquetador *Tree-Tagger*, um “etiquetador morfosintático independente de língua”²⁰⁵ (baseado em modelos estatísticos), disponibilizado no *site* do CEPRI²⁰⁶ por Berber Sardinha.

Para utilizar esse etiquetador, o primeiro passo é fazer o envio (*upload*) do *corpus* que se pretende etiquetar – no caso deste estudo, a lista de unidades de tradução. Para que o programa interpretasse cada unidade como uma frase, aumentando assim a chance de acerto na classificação de cada item das unidades plurivocabulares, foi necessário colocar um ponto final no fim de cada item da lista. Eis uma maneira de fazer isso automaticamente:

- copiar a coluna das unidades de tradução da tabela do Excel;
- colar num arquivo do Word;
- converter a tabela em texto, usando marcas de parágrafo como separadores (Fig. 46);
- introduzir um ponto final antes de cada marca de parágrafo (.^p) (vide Fig. 47) usando a ferramenta *Find & Replace* (Localizar & Substituir) do Word;
- salvar o resultado (Fig. 48) como um arquivo somente texto (.txt).

Figura 46: transformando coluna do Excel em texto

²⁰⁵ “language independent part-of-speech tagger”²⁰⁵

²⁰⁶ <http://www2.lael.pucsp.br/corpora/etiquetagem/index.html> . Além do português e do inglês, etiqueta textos em espanhol, alemão e italiano.

Figura 47: transformando marcas de parágrafo em ponto final

Figura 48: arquivo em .txt pronto para ser enviado para o etiquetador

Fiz, então, o *upload* desse arquivo na página apropriada do CEPRIL, anotei o código do corpus, fornecido pelo site, voltei à página do etiquetador, escolhi a língua apropriada (português), digitei esse código e solicitei a etiquetagem do arquivo (Fig. 49):

Figura 49: etiketador do CEPRIL

O resultado retornado pelo programa aparece na tela em três colunas: *palavra*, *etiqueta* e *lema* (palavra não flexionada). As etiquetas usadas pelo *Tree-Tagger* para o português (cujos parâmetros foram desenvolvidos por Pablo Gamallo Otero) são:

ADJ	Adjetivo
ADV	Advérbio
DET	Determinante
CARD	Número Cardinal / Ordinal
NOM	Nome Comum / Próprio
P	Pronome
PREP	Preposição
V	Verbo
I	Interjeição
VIRG	Separadores dentro da oração
SENT	Separadores de orações
PREP+DET	(exemplo: “do”, “das”, etc.)
V+P	(exemplo: “levou-me”, “disse-lhe”, etc.)

Não é possível colar o resultado fornecido pelo programa diretamente na tabela de estudo do Excel pois, para etiquetar, o programa *Tree-Tagger* faz a **itemização** do corpus, ou seja, coloca uma palavra em cada linha antes de atribuir-lhe a etiqueta correspondente (vide Fig. 50). Foi preciso processar o resultado para deixá-lo num formato adequado para colar na tabela – o que pode ser feito de várias formas, como, por exemplo, esta:

- copiar a tela de resultado do *Tree-Tagger*, sem as linhas de cabeçalho, e colar numa planilha de Excel;

- selecionar apenas a coluna do meio (que contém as etiquetas);
- colar num arquivo .txt (do Bloco de Notas), para eliminar qualquer tipo de formatação;
- copiar e colar num arquivo MS-Word para que se possa fazer a formatação dos dados usando a ferramenta Find& Replace (Procurar e Substituir).

	A	B	C	D	E	F	G	H	I	J	K
1654	abaixar	V	abaixar								
1655	o	DET									
1656	fogo	NOM	fogo								
		SENT									
1658	abaixe	V	abaixar								
1659	a	DET	a								
1660	massa	NOM	massa								
1661	.	SENT	.								
1663	o	DET	o								
1664	fogo	NOM	fogo								
1665	.	SENT	.								
1666	abobrinha	NOM	abobra								
1667	italiana	ADJ	italiano								
1668	.	SENT	.								
1669	abobrinhas	NOM	abobra								
1670	italianas	ADJ	italiano								
1671	.	SENT	.								
1672	abra	V	abrir								
1673	a	DET	a								
1674	massa	NOM	massa								
1675	.	SENT	.								

Figura 50: resultado do etiquetador colado em planilha do Excel

Em seguida, é preciso fazer com que os vários itens que compõem uma unidade plurivocabular sejam reagrupados numa linha só (vide quadro em destaque na Fig. 50). A etiqueta SENT, aplicada a todos os pontos finais acrescentados ao final de cada unidade, indica o limite entre elas. É preciso, portanto, eliminar as outras marcas de parágrafo, que foram acrescentadas apenas para itemizar as unidades analisadas, ou seja, todas aquelas que não forem seguidas da etiqueta SENT.

Primeiramente, aciona-se o botão que exhibe as marcas de parágrafo e formatação no texto do aplicativo Word. Em seguida, faz-se as seguintes substituições, usando a ferramenta *Find & Replace* do programa (vide Fig. 51 para ver a aparência do texto em cada uma das etapas):

- passo 1: trocar todas as ocorrências de “¶SENT¶” (*marca de parágrafo*SENT*marca de parágrafo*) por “→” (*marca de tabulação*) (vide coluna 2

da Figura 51). Digitar “^pSENT^p” na caixa Find/Procurar e “^t” (marca de tabulação) na caixa Replace/Substituir (sem as aspas);

- passo 2: trocar todas as marcas de parágrafo que sobraram por um sinal de + precedido e seguido de espaço em branco – para diferenciar da etiqueta “PRP+DET”, usada pelo etiquetador (resultado passo 2);
- passo 3: trocar as marcas de tabulação (→) por uma marca de parágrafo (^p).

etiquetador	resultado passo 1	resultado passo 2	resultado final
CARD¶	CARD¶	CARD.+NOM→DET.+	CARD.+NOM
NOM¶	NOM→DET¶	NOM→PRP+DET.+NO	DET.+NOM
SENT¶	NOM→PRP+DET¶	M→NOM→NOM→NOM	PRP+DET.+NOM
DET¶	NOM→NOM→NOM→	→V→V.+DET.+NOM→	NOM
NOM¶	NOM V→V¶	V.+DET.+NOM→V.+D	NOM
SENT¶	DET¶	ET.+NOM	NOM
PRP+DET¶	NOM→V¶	→V.+DET.+N	V
NOM¶	DET¶	OM→NOM.+ADJ→NO	V.+DET.+NOM
SENT¶	NOM→V¶	M.+ADJ→V.+DET.+N	V.+DET.+NOM
NOM¶	DET¶	OM→V.+DET.+NOM.+	V.+DET.+NOM
SENT¶	NOM→V¶	.PRP.+DET.+NOM→V.	V.+DET.+NOM
NOM¶	DET¶	+.DET.+NOM.+PRP.+	NOM.+ADJ
SENT¶	NOM→NOMv	DET.+NOM.+ADJ	NOM.+ADJ
V¶	ADJ→NOM¶		
SENT¶			
DET¶			

Figura 51: aparência do texto durante a formatação dos resultados do *Tree-tagger*

Após a realização desse procedimento, o resultado (coluna 4 acima) pôde ser copiado e colado diretamente na coluna reservada a ele na tabela Excel, como mostra a figura 52. Em seguida, contei quantos itens havia em cada unidade plurivocabular e coloquei esse valor numa nova coluna da planilha. Depois, passei à revisão e correção manual das etiquetas atribuídas pelo programa. Toda vez que uma etiqueta errada foi corrigida, aquela célula recebeu a cor laranja e a nova etiqueta foi escrita em letra minúscula. E toda vez que foi encontrada uma estrutura aberta (que pode ser completada por várias palavras de uma mesma categoria, como por exemplo “folha* de *NOM*” – em que *NOM* pode ser *alface*, *louro*, etc. –, “preaqueça o forno a *CARD*°C” – em que *CARD* por ser *180*, *200*, *250*, etc.), a célula recebeu a cor amarela:

unidade de tradução terminológica	Cat. Gram. itens	Cat. Gram. Geral	no. de elementos	contém pal. Chave	Column6
707 junte os ovos um a um	V + DET + NOM + card + PRP + card		6		
708 Karo	NOM		1		
709 kummel	NOM		1		
710 lagarto	NOM		1		
711 lagostin*	NOM		1		
712 lata de *	nom + PRP + nom		2		
713 lata grande de *	NOM + adj + PRP + nom		3		
714 lata pequena de *	NOM + adj + PRP + nom		3		
715 lata* de creme de leite	NOM + PRP + NOM + PRP + NOM		5		
716 lata* de leite condensado	NOM + PRP + NOM + adj		4		
717 lata* de tomate pelado	NOM + PRP + NOM + ADJ		4		
718 lata* de tomate pelado italiano	NOM + PRP + NOM + ADJ + ADJ		5		
719 leite condensado	NOM + adj		2		
720 leite de cabra	NOM + PRP + NOM		3		
721 leite de coco	NOM + PRP + NOM		3		
722 leite de soja	NOM + PRP + NOM		3		
723 leite de vaca	NOM + PRP + NOM		3		
724 leite morno	NOM + ADJ		2		
725 lentilha*	NOM		1		
726 lentilha* alaranjada*	NOM + ADJ		2		
727 lentilha* comu*	NOM + ADJ		2		
728 lentilha* du Puy	NOM + prp + NOM		3		
729 lentilha* francesa* du Puy	NOM + ADJ + prp + NOM		4		
730 lentilha* francesa* Puy	NOM + ADJ + NOM		3		
731 lentilha* Puy	NOM + NOM		2		
732 lev* a frigideira ao fogo novamente	V + DET + NOM + PRP&DET + NOM + ADV		&6		
733 lev* à geladeira por * hora*	v + PRP&DET + NOM + PRP + CARD + NOM		&6		

Figura 52: tabela com as unidades, categorias e nº. de elementos

Por fim, foi feita a classificação manual da função sintática geral das UTEs (coluna 3 da tabela – vide Fig. 8), usando a seguinte classificação:

adj	adjetivo isolado
adv	advérbio isolado
verb	verbo isolado
subs	substantivo isolado
SAdj	sintagma adjetival
SAdv	sintagma adverbial
SN	sintagma nominal (singular ou plural)
SNsg	sintagma nominal singular
SNpl	sintagma nominal plural
SV	sintagma verbal

4.2.2.2 Erros do etiquetador Tree-Tagger

Antes de iniciarmos a análise dos dados propriamente, cabe aqui fazer algumas observações sobre os erros cometidos pelo etiquetador Tree-Tagger na execução dessa tarefa. Um dos primeiros problemas enfrentados decorreu do fato de o programa ter interpretado erroneamente o ponto final acrescentado para marcar o fim de unidades terminadas em número (como, por ex., “cortada/o(s) em 4.”), como um ordinal (4. = 4º.). Isso fez com que todas as linhas da coluna, a partir desse ponto, não mais

coincidissem com a coluna de UTEs já que, em vez de duas etiquetas (CARD (“4”) SENT (“.”)), o etiquetador atribuiu apenas uma etiqueta ao trecho (“4.” = CARD – ordinal), eliminando SENT (o ponto final), como pode ser visto no trecho a seguir:

cortada	V	cortar
em	PRP	em
4.	DET	@ord@
cortadas	NOM	cortado
em	PRP	em
4.	NOM	@ord@
cortado	ADJ	cortado
em	PRP	em
4.	DET	@ord@

Assim, durante as substituições feitas para preparar o resultado para colá-lo na tabela Excel, a etiqueta não foi apagada, como aconteceu com as demais unidades plurivocabulares, fazendo com que a coluna ficasse, ao fim, com mais linhas que a coluna de UTEs. Após corrigir esse problema, todo o procedimento de formatar o resultado e colar novamente na planilha teve que ser refeito.

Também é possível perceber por essa pequena amostra que o programa cometeu vários erros de interpretação. O radical *cortad** foi etiquetado de três formas diferentes: *cortada* aparece como verbo, *cortadas* como substantivo e *cortado* como adjetivo. Fato semelhante ocorreu com o próprio “4.” não segmentado pelo programa: apesar de as ocorrências serem idênticas, ele as interpretou ora como determinante, ora como substantivo. Assim, julguei necessária uma correção das etiquetas atribuídas pelo Tree-Tagger não só para melhorar a acuidade dos resultados desta etapa do trabalho, mas também para identificar possíveis problemas quando da etiquetagem do *corpus* total para levantamento automático de sintagmas semelhantes e, principalmente, para melhorar o desempenho futuro do etiquetador em *corpora* da área de Culinária – fator que julgo da maior importância para a automatização da coleta de UTEs em *corpora* especializados.

Usando a ferramenta “Sort & Filter” (=Filtrar) do programa Excel (vide Fig. 53) e o concordanciador do programa WordSmith Tools (Scott, 1997), fiz uma análise geral dos problemas ocorridos na etiquetagem. Usei o filtro para exibir apenas as células em que havia ocorrido pelo menos um erro na etiquetagem (que havia sido realçado em laranja). Observou-se que, dentre as 1.294 UTEs em estudo, o programa errou pelo menos uma vez (ao menos um dos elementos foi etiquetado de forma errada) em 286 delas (22%).

Figura 53: uso da ferramenta *Sort & Filter* para localizar erros do etiquetador

Em seguida, para saber exatamente quais foram as categorias em que houve mais erros e em que porcentagem, copiei toda a coluna com os resultados da etiquetagem e coleí num arquivo do Bloco de Notas (.txt). Usando o concordanciador do programa *WordSmith Tools*, foram localizadas então as ocorrências corretas (em letras maiúsculas) e erradas (em minúsculas) de cada categoria. O resultado e as respectivas porcentagens de erro estão na tabela abaixo:

categoria	acertos	erros	porcentagem erro
adjetivo (ADJ/adj)	341	95	27,8%
substantivo (NOM/nom)	1.280	254	19,8%
cardinais (CARD/card)	11	2	18,2%
advérbio (ADV/adv)	45	6	11,3%
determinante (DET/det)	166	10	6,0%
verbo (V/v)	298	15	5,0%
preposição (PRP/prp)	570	6	1,0%
conjunções (CONJ/conj)	28	0	0%
TOTAL	2.739	388	14,1%

Tabela 7: número de ocorrências por categoria e porcentagem de erros

Embora o substantivo seja a categoria de maior ocorrência entre as UTEs, seja no número de etiquetas atribuídas correta ou incorretamente (1.534 oc. no total, 49,1%

de todas as unidades), em termos de porcentagem, a categoria dos adjetivos foi a que apresentou mais erros – 27,8% das ocorrências. Analisados os exemplos em que a etiquetagem foi equivocada observou-se que, para cada categoria parecia haver um tipo de palavra específico que dava margem à interpretação errônea do programa:

- i) no caso dos adjetivos, o programa interpretou muitos participípios (*picado, untada, forrada, dourado, etc.*) como verbo;
- ii) no caso dos substantivos, o problema foi principalmente com as palavras hifenizadas (*pimenta-do-reino, uva-passa, pão-duro*), mas também com os empréstimos (*nam pla, shoyu*) e os nomes terminados em “r” (*mixer, liquidificador, couve-flor*), todos interpretados como verbos;
- iii) no caso dos numerais, o programa interpretou “um” (*junte os ovos um a um*) como artigo indefinido;
- iv) no caso dos advérbios, o programa teve dificuldade com os que co-ocorreram com adjetivos (especialmente os participípios interpretados como verbos – *picado fino, ralado grosso, peneire junto*);
- v) no caso dos determinantes, o programa interpretou erroneamente alguns casos como pronomes pessoais (P) por ter errado a etiquetagem das palavras próximas – por exemplo, em *com o liquidificador ligado*, “o” foi interpretado como pronome porque “liquidificador” foi erroneamente etiquetado como verbo;
- vi) no caso dos verbos, o maior problema foram as formas imperativas, que o programa interpretou muitas vezes como substantivo ou adjetivo (*bata, pique*) – palavras que, na língua geral, podem ser interpretadas como tal, mas não em culinária (há exceções, como “leve”, que ocorre como adjetivo em alguns contextos, mas geralmente é verbo);
- vii) com as preposições, o problema parece ter sido em especial “a”, em expressões como *tempere a gosto* ou *a cada x minutos*, em que a preposição foi interpretada como artigo definido.

Uma possível justificativa para esses erros pode ser o fato de o *Tree-Tagger* ter sido treinado para etiquetar palavras inseridas em contextos e frases completas e estruturadas, e não listas de sintagmas, como as que são alvo deste estudo. Ou, mais provavelmente, esses erros se devam às discrepâncias entre as características do

corpus usado para “treinar” o etiquetador (constituído de textos da língua geral) e as especificidades que caracterizam os textos especializados da área Culinária.

Se não forem “treinados” especificamente para os textos de uma determinada área, os etiquetadores podem atribuir erroneamente, por exemplo, uma etiqueta de “verbo” a uma palavra hifenizada que não faz parte da sua lista de palavras previamente etiquetadas, pela simples presença do hífen (que ele identifica como pronominalização do verbo) – como parece ter ocorrido com “pimenta-do-reino” e “uva-passa”. Podem etiquetar todas as ocorrências de “clara” como adjetivo, porque essa é a função mais recorrente dessa palavra na língua geral, ao passo que, em Culinária é mais empregada com a função de substantivo – “clara” de ovo. Se fosse feita uma busca por padrões num *corpus* assim etiquetado, o resultado da busca por associações como, por exemplo, “substantivo + prep + substantivo”, “clara em neve” não constaria dos resultados.

Assim, é possível entrever que o sucesso dos métodos (semi-)automáticos de extração de padrões morfossintáticos dependerá, em muito, da precisão dos etiquetadores usados, e de sua adequação às características específicas da subárea em questão. Portanto, é necessário avaliar e, se necessário, revisar a anotação do *corpus* antes de empreender a busca por padrões.

4.2.2.3 Distribuição das UTEs por categoria morfossintática

Após a revisão das etiquetas, para facilitar a análise dos resultados os dados foram separados em diferentes planilhas do Excel, com base na categoria inicial de cada sintagma. A seguir, são apresentadas as tabelas que listam os padrões encontrados, o número de vezes que cada um ocorreu dentre as UTEs provenientes da memória de tradução e alguns exemplos ilustrativos. Na seqüência, são tecidas algumas considerações sobre os dados levantados nessa etapa do trabalho.

A. Substantivos e UTEs iniciadas por substantivo

PADRÃO	no. oc.	alguns exemplos
NOM	325	vagem, uva-passa, Karo, rúcula, shiitake, curry, óleo, farinha, sal, cebolinha

NOM + PRP + NOM	226	cúrcuma em pó, vinagre de arroz, colher de chá, espremedor de alho, espátula de borracha, fôrma de pudim, farelo de trigo, fava de baunilha, floretes de couve-flor, forminha de papel, martelo para bife
NOM + ADJ	158	couscous marroquino, filme plástico, massa folhada, coalhada seca, pimenta-do-reino branca
NOM + PRP + NOM + ADJ	34	azeite de oliva extravirgem ²⁰⁷ , creme de leite fresco, fôrma de bolo inglês, panela de fundo grosso, pão de fôrma integral, vinagre de vinho tinto
NOM + NOM	32	papel alumínio, batata bolinha, mostarda Dijon, chá mate, maçã Fuji, macarrão gravatinha
NOM + ADJ + PRP + NOM	24	superfície polvilhada com farinha, atum conservado em água, gelatina incolor sem sabor
NOM + PRP + NOM + PRP + NOM	16	água de flor de laranjeira, cubo de caldo de carne, pedaço de canela em pau, vidro de leite de coco
NOM + ADJ + ADJ	15	massa folhada congelada, pão francês amanhecido
NOM + QUOTE + NOM + QUOTE	6	colher (chá), xíc. (chá), colh (sopa), xícara (chá)
NOM + PRP&DET + NOM	5	chocolate ao leite, frutos do mar, pimenta da jamaica
NOM + PRP + CARD + NOM	3	mistura de 7 grãos, suco de 2 laranjas, suco de um limão
NOM + ADV + ADJ	3	xícara bem cheia, azeite extra virgem, chocolate meio amargo
NOM + ADJ + PRP&DET + NOM	3	fôrma furada no centro, noz-moscada ralada na hora, pimenta-do-reino moída na hora
NOM + PRP + V + NOM	2	máquina de fazer macarrão, máquina de moer carne
NOM + PRP + V	2	faca para trincar, prato de servir
NOM + NOM + ADJ	2	mostarda Dijon al'ancienne, mostarda Dijon ancienne
NOM + CONJ + NOM + PRP + NOM	2	coxa e sobrecoxa de frango, coxa e sobre-coxa de frango
NOM + CONJ + NOM	2	coxa e sobrecoxa, coxa e sobre-coxa
NOM + ADJ + V + PRP&DET + NOM	2	pimenta-do-reino preta moída na hora, pimenta-do-reino branca moída na hora
NOM + ADJ + PRP + NOM + ADJ	2	queijo parmesão em lascas finas, queijo parmesão em lascas grossas
NOM + ADJ + CONJ + ADJ	2	ervilha fresca ou congelada, ervilha congelada ou fresca
NOM + PRP + NOM + ADJ	1	pistache sem casca salgado
NOM + PRP + NOM + PRP&DET + NOM	1	fôrma com buraco no meio

²⁰⁷ como pode ser observado nessa UTE, se a comparamos com a ocorrência “azeite extra virgem” no padrão NOM + ADV + ADJ que está 8 linhas depois, para o programa, palavra é aquilo que tem um espaço em branco antes e um depois, ou seja, “extravirgem” será etiquetado como ADJ, mas “extra virgem” como ADV + ADJ (ainda que as duas palavras, juntas, tenham a função de adjetivo).

NOM + PRP + NOM + PRP + NOM + NOM	1	folha de massa para rolinho primavera
NOM + PRP + NOM + PRP + NOM + ADJ	1	forma para quiche de fundo removível
NOM + PRP + NOM + NOM	1	massa para rolinho primavera
NOM + PRP + NOM + ADV + ADJ	1	azeite de oliva extra virgem
NOM + PRP + NOM + ADJ + CONJ + ADJ	1	castanha de caju torrada e salgada
NOM + PRP + NOM + ADJ + ADJ	1	lata de tomate pelado italiano
NOM + PR + V + ADV	1	queijo que derrete bem
NOM + NOM + QUOTE + NOM + PRP + NOM + QUOTE	1	nam pla (molho de peixe)
NOM + NOM + ADJ + PRP + NOM	1	mostarda Dijon ancienne com grãos
NOM + DET + NOM + ADJ	1	gema de ovo caipira
NOM + ADJ + VIRG + ADJ + CONJ + PRP + NOM	1	amendoim torrado, salgado e sem pele
NOM + ADJ + NOM + NOM	1	cogumelo fresco tipo Paris
NOM + ADJ + NOM	1	lentilha francesa Puy
NOM + ADJ + ADV + ADJ + PRP&DET + NOM	1	açúcar mascavo bem apertado na xícara
NOM + ADJ + ADV + ADJ + CONJ + ADJ	1	arroz branco bem lavado e escorrido
NOM + ADJ + ADJ + CONJ + ADJ	1	arroz branco lavado e escorrido
NOM + ADJ + ADJ + ADJ	1	massa folhada congelada laminada
TOTAL:	886	

Tabela 8: substantivos e UTEs iniciados por substantivo

B. Verbos e UTEs iniciadas por verbo

PADRÃO	no. oc.	exemplos
V	89	amasse, caramelizar, cozinhe, descongele, escale, escorra, gratinar, mexendo, seque
V + DET + NOM	26	acertar o sal, corte uma tampinha, descasque o ovo, trabalhando a massa
V + PRP + NOM	17	adoce a gosto, asse em banho-maria, deixe de molho, polvilhe com parmesão, regue com azeite
V + PRP&DET + NOM	16	cozinhe no vapor, leve ao freezer, doure na beirada, retire da fôrma, soque num pilão
V + PRP + DET + NOM	12	adoce com o açúcar, regue com o azeite, retire com a escumadeira, salpique com a salsinha
V + PRP + NOM + ADJ	10	cubra com filme plástico, passe por água fria, escorra sobre papel absorvente, cozinhe em fogo alto
V + V	9	deixe esfriar, deixe descansar, espere ferver
V + ADV	9	misture delicadamente, peneire junto, aqueça bem
V + DET + NOM + PRP&DET + NOM	7	faça um buraco no meio, abra um buraco no centro, retire os talos da alface
V + PRP&DET + NOM + PRP + CARD + NOM	4	leve à geladeira por 30 minutos, leve ao freezer por 2 horas, leve ao congelador por 40 minutos

V + PRP + DET + NOM + ADJ	3	polvilhe com o queijo ralado, polvilhe com o queijo parmesão, guarde em um pote fechado
V + DET + NOM + ADJ + PRP&DET + ADJ	2	junte a calda quente aos poucos, junte a calda quente aos pouquinhos
V + DET + NOM + CONJ + DET + NOM	2	acerte o sal e a pimenta, acerte o sal e a pimenta-do-reino
V + DET + NOM + PRP&DET + NOM + ADV	2	leve a frigideira ao fogo novamente, leve a panela ao fogo novamente
V + NOM	2	espalhe manteiga, espete cravinhos
V + PRP + NOM + PRP + ADV + PRP + DET + NOM + PRP&DET + NOM	2	regando de vez em quando com a gordura da assadeira, regando de vez em quando com o líquido do cozimento
V + PRP + NOM + PRP + NOM	2	cubra com filme de PVC, escorra sobre toalhas de papel
V + PRP&DET + NOM + ADJ	2	rale no ralo grosso, guarde num pote fechado
V + PRP&DET + NOM + ADV	2	leve ao fogo novamente, leve ao forno novamente
V + V + NOM	2	deixe levantar fervura, deixe tomar gosto
V + ADV + ADJ	1	pique bem fininho
V + ADV + DET + NOM	1	doure ligeiramente a cebola
V + CONJ + V	1	unte e polvilhe
V + CONJ + V + PRP + NOM	1	unte e polvilhe com farinha
V + DET + NOM + ADJ + PRP + DET + NOM	1	deixando um espaço livre entre os pães
V + DET + NOM + ADJ + PRP + P	1	deixando um espaço livre entre elas
V + DET + NOM + CARD + PRP + CARD	1	junte os ovos um a um
V + DET + NOM + PR + V	1	retire a espuma que subir
V + DET + NOM + PR + V + PRP + DET + NOM	1	retire a espuma que subir com uma escumadeira
V + DET + NOM + PRP + CARD + NOM	1	aqueça o forno a 400C°
V + DET + NOM + PRP + DET + NOM	1	abra a massa com o rolo
V + DET + NOM + PRP + DET + NOM + ADJ	1	abra a massa sobre uma superfície enfarinhada
V + DET + NOM + PRP + DET + NOM + ADJ + PRP + NOM	1	abra a massa sobre uma superfície polvilhada com farinha
V + DET + NOM + PRP + NOM	1	aqueça um fio de azeite
V + DET + NOM + PRP + V	1	bata as claras até espumar
V + PRP + ADJ + CARD + NOM	1	regando a cada 15 minutos
V + PRP + DET + NOM + PRP + NOM	1	adoce com um pouco de açúcar
V + PRP + NOM + CONJ + NOM	1	polvilhe com açúcar e canela
V + PRP + NOM + NOM	1	escorra sobre papel toalha
V + PRP + NOM + PRP + ADV	1	mexendo de vez em quando
V + PRP + NOM + PRP + ADV + PRP + DET + NOM	1	regando de vez em quando com a marinada
V + PRP + V	1	mexendo sem parar
V + PRP&DET + NOM + PRP + NOM	1	cozinhe na panela de pressão
V + PRP + ADJ + CARD + NOM	1	regue a cada 10 minutos
V + V + PRP&DET + ADJ	1	vá juntando aos poucos
TOTAL:	244	

Tabela 9: Verbos e UTEs iniciadas por verbo
C. Adjetivos e UTEs iniciadas por adjetivos

PADRÃO	no. oc.	exemplos
ADJ	54	antiaderente, clarificada, descongelado, fatiadas, embebidos, grelhados, gelatinosa, limpo
ADJ + CONJ + ADJ	4	lisa e elástica, lisa e homogênea, lisa e macia, lisa e uniforme
ADJ + ADV	2	picado fino, ralado grosso
ADJ + PRP + ADV	2	dourada por cima, rosado por dentro
ADJ + PRP + DET + NOM	2	forrada com um pano, regada com o molho
ADJ + PRP + NOM	2	regado com azeite, forrado com papel-toalha
ADJ + PRP&DET + NOM	2	dourado na beirada, dourado nas bordas
ADJ + ADV + ADJ	1	picado bem fininho
ADJ + PRP + ADV + NOM	1	fritos em bastante óleo
ADJ + PRP + CARD	1	cortados em 4
ADJ + PRP&DET + NOM + ADJ	1	ralados no ralo grosso
TOTAL:	72	

Tabela 10: adjetivos e UTEs iniciadas por adjetivos

D. UTEs iniciadas por preposições

PADRÃO	no. oc.	exemplos
PRP + NOM	14	a gosto, com casca, em cunhas, em cubos, em palitinhos, em bocados, em cubinhos
PRP + DET + NOM + ADJ	6	com a batedeira ligada, com a panela destampada, com a panela semitampada
PRP + NOM + ADJ	6	em fatias finas, em rodela grossas, em tiras finas
PRP + DET + NOM	4	com a casca, com a tampa, sem a casca, sem a tampa
PRP + NOM + ADV + ADJ	4	em tiras bem fininhas, em fatias bem grossas
PRP + V + ADJ	4	até ficar crocante, até ficar dourado, até ficar macio
PRP&DET + NOM	4	à juliana, aos bocados, ao ponto, no ponto
PRP + PR + V + ADJ	3	até que esteja pronto, até que estejam gratinadas
PRP + PR + V + PRP&DET + NOM	3	até que solte das mãos, até que esteja no ponto, até que desgrude das mãos
PRP + PR + V	2	até que murchem, até que evapore
PRP + PR + V + ADJ + CONJ + ADJ	2	até que esteja macio e dourado, até que esteja lisa e brilhante
PRP + PR + V + DET + NOM + P + V + ADJ	2	até que enfiando um palito ele saia limpo, até que enfiando uma faca ela saia seca
PRP + V	2	até ferver, até gratinarem

PRP&DET + ADJ	2	aos poucos, aos pouquinhos
PRP + DET + NOM + ADJ + PRP + ADJ	1	com a boca virada para baixo
PRP + DET + NOM + PRP + NOM + ADJ	1	com a batedeira em velocidade baixa
PRP + DET + NOM + PRP + V	1	até a hora de servir
PRP + PRP&DET + ADV + CARD + NOM + PRP + NOM	1	com pelo menos 12 horas de antecedência
PRP + V + CONJ + V + DET + NOM + PRP + DET + NOM	1	até encorpar e cobrir o dorso de uma colher
PRP + V + CONJ + V + DET + NOM + PRP&DET + NOM	1	até encorpar e cobrir o dorso da colher
PRP + V + DET + NOM + PRP&DET + NOM	1	até aparecer o fundo da panela
PRP + V + NOM	1	até levantar fervura
PRP + V + PRP + NOM	1	até dobrar de volume
PRP + V + PRP + V	1	até começar a secar
PRP + V + PRP + V + DET + NOM + PRP&DET + NOM	1	até começar a aparecer o fundo da panela
PRP&DET + ADV + NOM	1	num só pedaço
PRP&DET + NOM + ADV	1	num pedaço só
PRP&DET + NOM + PRP&DET + NOM	1	no sentido do comprimento
TOTAL:	72	

Tabela 11: UTEs iniciadas por preposições

E. Advérbios e UTEs iniciadas por advérbios

PADRÃO	no. oc.	exemplos
ADV	2	aproximadamente, grosseiramente
ADV + ADJ	8	bem picado, grosseiramente picados, não untada
ADV + DET + NOM + ADJ	2	só a parte branca, somente a parte branca
ADV + V	2	sem untar ²⁰⁸ , sempre mexendo
ADV + V + PRP + V ²⁰⁹	2	sem parar de bater, sem parar de mexer
TOTAL:	16	

Tabela 12: advérbios e UTEs iniciadas por advérbios

²⁰⁸ numa etapa muito posterior apenas percebi esse erro de classificação – sem é preposição, e não advérbio. A categoria “prp + v” está prevista (vai recuperar “sem untar”), mas a categoria “adv + v + prp + v” (próxima linha da coluna) poderia ter sido eliminada, conforme se constatou posteriormente nas buscas feitas no corpus.

²⁰⁹ vide nota anterior.

F. UTEs iniciadas por conjunções

PADRÃO	no. oc.	exemplos
CONJ + DET + NOM + V	1	enquanto o forno aquece
CONJ + P	1	enquanto isso
CONJ + V	1	quando ferver
CONJ + V + adj	1	quando estiver pronto
TOTAL:	4	

Tabela 13: UTEs iniciadas por conjunções

G. UTEs iniciadas por outras categorias

PADRÃO	no. oc.	exemplos
DET + NOM + CONJ + DET + NOM + PRP + nom	1	as raspas e o suco de 1 limão
PR + V + V + PRP&DET + NOM	1	que possa ir ao forno

Tabela 14: UTEs iniciadas por outras categorias

4.2.2.4 Análise dos dados e algumas considerações preliminares

De modo geral, os resultados parecem confirmar o esperado – a grande maioria das UTEs constitui-se de ou começa com um substantivo (68%), e em seguida vêm as unidades iniciadas por verbo (18,9%). Mas as ocorrências de UTEs iniciadas por adjetivos (5,5%), preposições (5,5%) e advérbios (1,2%) – além de outras categorias menos expressivas (conjunção, pronome relativo e determinante – que, juntas, somam cerca de 1%) respondem por mais de 10% do total de UTEs. Isso já é um indício de que, em consonância com a hipótese central do presente trabalho, as unidades de tradução especializadas de uma dada área técnico-científica parecem ir além das categorias tradicionais (nome e verbo), compreendendo também estruturas textuais típicas da comunicação especializada da área que, quer contenham termos ou representem nós conceituais de um conhecimento estruturado, quer não, são de suma importância para um público-alvo específico: o tradutor, que é, acima de tudo, um produtor de textos.

Quanto à função sintática das UTEs como um todo, o número de formações sintagmáticas (827 oc.) é praticamente o dobro do número de categorias que ocorrem

sozinhas (467 oc. – 325 substantivos, 86 verbos, 55 adjetivos e 2 advérbios). Levando-se em conta que esse glossário foi construído praticamente de forma aleatória, conforme essas unidades coletadas se mostrassem necessárias no trabalho de tradução, poderíamos considerar esse resultado como uma sugestão de que as formações sintagmáticas apresentam maior dificuldade para o tradutor – razão pela qual devem constar em grande número nos dicionários voltados para esse público.

Os sintagmas nominais (SN) predominam, com 563 ocorrências – sendo que 213 deles ocorrem preferencialmente no singular (SNsg), como *salsa crespa* e *pimenta-do-reino branca*, e 3 preferencialmente no plural: *as raspas* e *o suco de um limão, algas marinhas e frutas vermelhas*. Em seguida vêm os sintagmas verbais (SV), com 157 ocorrências e, em 3º. lugar, corroborando mais uma vez a hipótese de que não só os substantivos, verbos e adjetivos importam para o tradutor, vem a categoria dos sintagmas adverbiais (SAdv), com 74 unidades. Por último, vêm os sintagmas adjetivais (SAdj), com 33 ocorrências.

A meu ver, a alta ocorrência de sintagmas adverbiais dentre as UTEs estudadas parece refletir uma particularidade do discurso culinário que, por seu caráter processual e prático, preza a minúcia, o detalhe e a precisão para a correta caracterização dos processos e produtos envolvidos – função cumprida pelos advérbios. Parece óbvio, portanto, que essa categoria, em consonância com os achados de Moraes (2004) para a língua geral e a lexicografia, não pode mais ser desprezada ou relegada a segundo plano na compilação de dicionários.

Observemos agora as UTEs quanto ao número de elementos que as compõem. O gráfico da figura 54 representa a proporção entre o tamanho das unidades encontradas (eixo x) e o número de vezes que cada uma ocorreu (eixo y) – 2 elementos (258 oc.), 3 elementos (344 oc.), 4 elementos (126 oc.), 5 elementos (56 oc.), 6 elementos (26), 7 elementos (4 oc.), 8 elementos (6 oc.), 9 elementos (2 oc.) e 10 elementos (2 oc.). Cada cor representa um tipo de sintagma.

No gráfico não estão representadas as 463 ocorrências de unidades com apenas 1 elemento (substantivos – 325 oc., verbos – 86 oc., adjetivos – 55 oc., advérbios – 2 oc.) pois, além de sua alta ocorrência prejudicar a representação dos dados no gráfico, o principal interesse do presente trabalho, como já foi dito, está nas unidades plurivocabulares.

Figura 54: relação entre número de elementos e categoria gramatical

A única classe gramatical que ocorreu em UTEs de todos os tamanhos foi o verbo – que considero outro reflexo das características do gênero textual “receita” (Teixeira, 2004: 189), que é *operativo* (Nord, 1997: 38) / *instrucional* (Colina, 1997: 336), ou seja, focaliza os processos, usando os verbos e sintagmas verbais para representá-los lingüisticamente. Os advérbios também apresentaram uma distribuição uniforme, estando ausentes apenas nas UTEs com 10 elementos – isso, mais uma vez, confirma sua importância na caracterização do perfil morfo-sintático-semântico da Culinária.

Quando da etiquetagem e contagem do número de elementos de cada UTE, foram marcados os casos em que a combinação da preposição com um artigo (PRP&DET) constituía um dos elementos da unidade (por exemplo, *aos poucos*, *frutos do mar*, *cozinhe na pressão*, *deixe na geladeira*, *retire do fogo*) – foram 64 ocorrências. A princípio, essa informação parece ter pouca importância, especialmente se, momentaneamente, a faceta bilíngüe do glossário tenha sido desprezada. Mas, certamente, será muito útil para a produção textual do falante cuja língua nativa não é o português. Além disso, é possível que uma busca por estruturas sintáticas que contenham PRP&DET precedido de verbo, por exemplo, retorne outras unidades importantes para o tradutor. Outro caractere que apareceu na lista e que pode ser útil na recuperação de novas UTEs de um *corpus* de Culinária são os parênteses – como, por exemplo, em *colher (sopa)*, *xíc. (chá)*, *endro (dill)* e *nam pla (molho de peixe)*.

Enfim, uma vez que minha lista não foi construída seguindo critérios necessários para que fosse representativa, me propus a remediar minimamente essa falha respondendo às seguintes questões:

- a UTE contém alguma palavra-chave?
- qual a frequência da UTE no *corpus* total?

Em posse das respostas a essas perguntas, os resultados obtidos até aqui foram reconsiderados para que, então, as estruturas mais produtivas, a serem usadas na extração automatizada, fossem identificadas.

4.2.3 UTEs e palavras-chave

Para obter as palavras-chave de um dado *corpus* é necessário comparar uma lista de todas as palavras nele contidas, acompanhadas de suas respectivas frequências, com a lista de palavras (do mesmo tipo) do *corpus* de referência, para que os cálculos comparativos possam ser efetuados e a significância estatística determinada.

O *WordSmith Tools*® (Scott, 1997)²¹⁰, como já vimos no Capítulo III (item 3.3.4) é um dos programas com os quais é possível compilar essas listas automaticamente (ferramenta *WordLists*) e, a partir delas, gerar a lista de palavras-chave do *corpus* de estudo usando a ferramenta *Keywords* (Fig. 55 e 56). Vide, por exemplo, na figura 55, que a palavra de maior ocorrência no *corpus* de estudo é a preposição “de” (ocorre 125.837 vezes, ou seja, representa 8,27% de todo o corpus). No entanto, a mesma palavra cai para a posição 32 na lista de palavras-chave (Fig. 56). Isso significa que, apesar de sua ocorrência ser grande também no *corpus* de referência (4,66% do corpus), sua presença no *corpus* especializado é estatisticamente significativa – “de” é uma **palavra-chave** da culinária. O mesmo não acontece, por exemplo, com a preposição “em”.

Nas primeiras posições da lista de palavras-chave é possível identificar o vocabulário que tipifica as receitas culinárias. Muitas dessas palavras-chave poderiam ser distribuídas em um sistema conceptual estruturado da área, por exemplo: unidades de medida (“colheres”, “xícara”), ingredientes (“sal”, “manteiga”, “farinha”), processos

²¹⁰ para mais informações sobre o programa, vide: <http://www.lexically.net/wordsmith/>. Para informações sobre como gerar listas de palavras (*wordlists*) e as listas de palavras-chave (*keyword lists*) usando esse programa, vide capítulo 3 em Berber Sardinha (2004) e/ou apostila *Como usar o WordSmith Tools* (Teixeira, 2003).

(“coloque”, “misture”, “retire”), equipamentos (“panela”, “forno”), produtos (“molho”, “massa”). Mas outras – como o próprio “de”, “minutos”, “modo” – não, pois são itens lexicais que participam de estruturas textuais maiores. Embora ocorram com grande frequência em textos da área, essas estruturas nem sempre contêm um termo (como “**de** vez em quando”) e, não raro, são bastante longas (“leve ao **forno** preaquecido a 200oC por cerca de 30 minutos”). Por isso, não costumam ser incluídas como entrada em dicionários técnicos.

N	Word	Freq	% Lemmas
1	DE	125.837	8.27
2	E	67.127	4.41
3	O	46.893	3.08
4	A	42.835	2.82
5	COM	29.401	1.93
6	EM	27.916	1.84
7	POR	15.447	1.02
8	SOPA	15.115	0.99
9	UMA	13.856	0.91
10	SAL	12.844	0.84
11	AS	11.989	0.79
12	ATÉ	11.829	0.78
13	CHÁ	11.226	0.74
14	OS	11.048	0.73
15	PARA	10.729	0.71
16	MINUTOS	10.549	0.69
17	PREPARO	10.209	0.67
18	DO	9.762	0.64
19	COLHERES	9.705	0.64
20	UM	9.545	0.63
21	ÁGUA	9.221	0.61
22	INGREDIENTES	9.134	0.60
23	XICARA	8.994	0.59
24	COLHER	8.862	0.58
25	AO	8.465	0.56
26	COLOQUE	8.216	0.54
27	FOGO	8.179	0.54
28	OU	7.730	0.51
29	LEITE	7.529	0.50
30	BEM	7.499	0.49
31	NA	7.333	0.48
32	MANTEIGA	7.015	0.46
33	G	6.705	0.44
34	DEIXE	6.588	0.43
35	MODO	6.464	0.43
36	AÇÚCAR	6.422	0.42
37	MASSA	6.414	0.42
38	LEVE	6.140	0.40

Figura 55: lista de palavras (Wordlist) do *corpus* de culinária – português

N	WORD	FREQ	TESEPOTT LST %	FREQ	BANCOP-1 LST %	KEYNESS	P
1	SOPA	15.115	0.99	3.192	135.226,5	0.000000	
2	SAL	12.844	0.84	3.621	111.950,8	0.000000	
3	CHÁ	11.226	0.74	4.202	94.943,0	0.000000	
4	COLHERES	9.705	0.64	945	91.274,1	0.000000	
5	PREPARO	10.209	0.67	2.758	89.332,3	0.000000	
6	XÍCARA	8.994	0.59	1.004	83.979,8	0.000000	
7	INGREDIENTES	9.134	0.60	3.728	76.457,3	0.000000	
8	COLHER	8.862	0.58	3.405	74.710,3	0.000000	
9	COLOQUE	8.216	0.54	2.638	70.650,1	0.000000	
10	MANTEIGA	7.015	0.46	1.958	61.176,9	0.000000	
11	MINUTOS	10.549	0.69	35.985	56.795,9	0.000000	0.02
12	MISTURE	5.640	0.37	643	52.588,9	0.000000	
13	RETIRE	5.594	0.37	908	51.020,8	0.000000	
14	DEIXE	6.588	0.43	4.953	50.567,2	0.000000	
15	PANELA	5.760	0.38	1.518	50.503,4	0.000000	
16	FOGO	8.179	0.54	19.577	48.881,7	0.000000	
17	FORNO	5.739	0.38	2.077	48.702,5	0.000000	
18	FARINHA	5.910	0.39	3.091	47.906,8	0.000000	
19	JUNTE	4.999	0.33	680	46.127,4	0.000000	
20	AZEITE	5.237	0.34	1.392	45.879,2	0.000000	
21	ÁGUA	9.221	0.61	41.810	45.098,7	0.000000	0.02
22	MOLHO	5.766	0.38	3.973	44.880,4	0.000000	
23	LEITE	7.529	0.50	20.525	43.373,6	0.000000	
24	TEMPERATURA	6.122	0.40	7.386	43.068,9	0.000000	
25	CEBOLA	5.046	0.33	1.794	42.904,6	0.000000	
26	AÇÚCAR	6.422	0.42	9.931	42.817,5	0.000000	
27	G	6.705	0.44	13.006	42.341,7	0.000000	
28	CREME	5.063	0.33	2.551	41.246,7	0.000000	
29	LEVE	6.140	0.40	9.425	41.005,2	0.000000	
30	ACRESCENTE	4.461	0.29	734	40.645,1	0.000000	
31	MASSA	6.414	0.42	12.975	40.073,8	0.000000	4.66
32	DE	125.837	8.27	10.729.884	36.478,4	0.000000	
33	PIMENTA-DO-REINO	3.557	0.23	0	35.773,0	0.000000	
34	ALHO	4.236	0.28	1.731	35.439,3	0.000000	
35	SIRVA	3.870	0.25	1.131	33.589,5	0.000000	
36	DIFICULDADE	5.640	0.37	14.143	33.266,4	0.000000	
37	MODO	6.464	0.43	26.318	32.815,5	0.000000	0.01
38	QUENTE	4.765	0.31	6.554	32.604,2	0.000000	

Figura 56: lista de palavras-chave (keywords) do *corpus* de português

Foi feita uma lista de palavras-chave do *corpus* de culinária de português usando $p = 0,001$ (Fig. 57) – i.e., em termos desta pesquisa, há uma chance em 1.000 de que as palavras da lista resultante não sejam chave (ou seja, de que a significância de sua ocorrência nesse conjunto de textos se deva ao acaso). O programa Keywords retornou 4.173 palavras do *corpus* de português que ocorrem tipicamente em textos desse tipo e, portanto, caracterizam o léxico da área.

Figura 57: janela em que se escolhe o p a ser usado na *Keyword List*

4.2.3.1 Verificando a presença de palavras-chave nas UTEs

Retomando as perguntas que se esperava responder nesta fase do estudo, o passo seguinte foi comparar a lista de UTEs proveniente da memória de tradução, palavra a palavra, com essa lista de palavras-chave gerada pelo utilitário Keyword, para saber quais delas eram chave no *corpus* de Culinária de português. Para tanto, foram usadas rotinas escritas na linguagem de programação Shell, do sistema operacional Unix²¹¹.

Essa linguagem permite fazer, entre muitas outras coisas, diferentes tipos de busca, substituição, ordenação e contagem de palavras e elementos do texto de forma automática. Quem possui um computador com sistema operacional diferente do Windows pode instalar um emulador de ambiente Unix, o *Cygwin*²¹² – ele fará com que o computador se comporte como se utilizasse o sistema operacional Unix, mas sem interferir com o ambiente Windows. Na figura a seguir é possível ver a tela do *Cygwin* (tela de fundo preto à esquerda) rodando no sistema operacional Windows Vista:

Figura 58: emulador de Unix Cygwin rodando em ambiente Windows

²¹¹ O co-orientador deste trabalho, Prof. Dr. Tony Berber Sardinha, da PUC - São Paulo, nos auxiliou com seus conhecimentos computacionais nesta etapa da pesquisa.

²¹² para maiores informações de como baixar e instalar o Cygwin (disponível em: www.cygwin.com), vide Berber Sardinha (2004: 54)

Para escrever uma rotina em Shell é preciso conhecer os utilitários entendidos pelo sistema *Unix*, por exemplo: *sed*, *grep*, *tr*, *sort*, etc., cujo manual, “man”, consta do próprio Cygwin, bastando digitar na tela do programa, após o *prompt* de comando (um cifrão - \$) *man nome-do-comando*; por exemplo, *man grep*. Lá o usuário encontra uma lista dos comandos desse utilitário e o que cada um faz. Não seria pertinente detalhar aqui o funcionamento de todos esses utilitários – vou ater-me apenas à explicação das rotinas utilizadas para efetuar as pesquisas deste trabalho.

O *Shell* trabalha com arquivos do tipo texto (inclusive as rotinas são escritas também num arquivo do Bloco de Notas). Portanto, a primeira coisa que fiz foi copiar e colar a coluna “WORD” da tela de resultados do Keyword, que contém as palavras-chave do corpus, num arquivo somente texto (Fig. 59). Apesar da aparência de “texto corrido” mostrada na tela, há uma marca de quebra de linha entre cada uma das palavras-chave.

Em seguida, preparei a lista de UTEs retiradas da memória de tradução para que pudesse ser lida pelo Shell, ou seja, a exemplo do que foi feito para o uso do etiquetador Tree-Tagger, coloquei uma palavra em cada linha (itemização), usando o ponto final para delimitar as UTEs (vide Fig. 60). Além disso, apaguei as partes variáveis das palavras, para que o programa, ao buscar, por exemplo, *salpi.** na lista de palavras-chave, retornasse uma resposta positiva quer encontrasse *salpicar*, *salpique*, *salpicado*, etc. Por fim, escrevi a rotina, que foi testada e reformulada várias vezes. A versão final pode ser vista na Fig. 61 (as explicações para cada linha de comando estão na seqüência)

Figura 59: palavras-chave em .txt

Figura 60: UTEs itemizadas

Figura 61: rotina em Shell para procurar palavras-chave

#!/bin/sh esta linha inicia um programa escrito em Shell

while estabelece um loop de leitura (**read**), ou seja, repete os comandos que vão de **do** a **done** até chegar ao fim da lista de **UTES** – o programa lê linha por linha desta lista (Figura 60) e executa os comandos

freq=\$ significa "a variável *freq* é o resultado dos comandos que estão entre parênteses"

grep significa "procure"; **-c** significa "conte quantas vezes encontrou e me retorne apenas esse número"; **-i** significa "ignore se as letras são maiúsculas ou minúsculas"

^\$UTES **2kwrds_p0.001.txt** - significa "faça isso (a contagem, ignorando se as letras são maiúsculas ou minúsculas), dentro do arquivo a seguir (*2kwrds_p0.001.txt*, o arquivo em que armazenei as palavras-chave com $P > 0.001$), buscando uma linha (=palavra) de cada vez do arquivo cujo nome aparece depois de **done** (*UTES.item.txt*); o circunflexo **^** faz com que a busca seja feita apenas no início da palavra

echo significa "imprima / escreva"; **-e** significa que vou usar um símbolo especial (**\t**, que significa 'marca de tabulação') que não deve ser interpretado literalmente pelo programa, ou seja, não é para o programa escrever na tela "palavra \t frequência", mas "palavra frequência";

\$UTES e **\$freq** são, respectivamente, as variáveis criadas para armazenar a palavra de busca (UTES) e a frequência da mesma no corpus, ambas preenchidas com valores advindos de cada vez que o programa executa o *loop*.

done significa que o loop termina ali

< significa "de onde o programa deve retirar as linhas" - neste caso, para preencher a variável UTEs (*UTES.item.txt* - arquivo com todas as UTEs já itemizadas)

> significa "onde o programa deve imprimir / escrever o resultado".

Para que o Cygwin execute essa rotina é preciso digitar na tela²¹³:

\$ sh **nome-da-rotina.sh**

O resultado fornecido pelo programa (o arquivo *freq.kw.PO.txt*, que contém uma coluna com as palavras procuradas e outra com suas respectivas freqüências no corpus), a exemplo do que ocorreu quando o etiquetador foi usado, não pode ser colado diretamente na tabela *Excel*. Por isso, executei um procedimento semelhante ao descrito no item 4.2.2 para formatar o resultado antes de colá-lo na tabela. Em seguida, troquei a seqüência de números obtida (as freqüências dos itens de cada UTE) por uma resposta à pergunta: a UTE contém alguma palavra-chave? (0 = não, >0 = sim). Por exemplo.:

UTE:	acerte	o	sal	e	a	pimenta	
é chave?	sim	não	sim	sim	não	sim	
resultado:	1	0	1	1	0	1	= 4 (sim)

4.2.3.2 Análise dos dados e algumas considerações preliminares

Das 1.294 UTEs estudadas, 1.181 contêm pelo menos uma palavra-chave. Há 113 que não possuem nenhuma palavra chave: 6 compostas por 2 elementos e 107 com apenas um elemento (vide lista a seguir) – isso não significa que as palavras não ocorram no *corpus* (o que será verificado no passo a seguir), mas sim que sua freqüência isolada não é estatisticamente significativa se comparada à sua freqüência num *corpus* de língua geral (no caso desta pesquisa, o Banco de Português).

²¹³ foi observado por alguns usuários de versões mais recentes do Windows Office que o Bloco de Notas acrescenta uma inadvertida marca de parágrafo ('r') ao final de todas as linhas – caso em que o comando digitado deve ser: `tr -d "\r" < nome-da-rotina.sh | sh` (para eliminar esses retornos indesejáveis, que causam erro na execução da rotina).

(2 elementos = 6)

- à Juliana
- aos bocados
- aos pouquinhos
- balança digital
- enquanto isso
- feno grego

(1 elemento = 107)

- alecrim-de-cheiro
- alface-lisa
- alface-romana
- alfavaca-do-campo
- almofariz.*
- aneto
- anis-doce
- arroz-selvagem
- badiana
- badiana-da-china
- badiana-de-cheiro
- balança.*
- banana.*-passa.*
- batata.*-bolinha
- batata.*-rosada.*
- batata.*-roxa.*
- branque.*
- cana-do-brejo
- capim-de-cheiro
- capim-limão
- cariz
- cebola.*-roxa.*
- cebolinha-francesa
- cebolinha-verde
- cerefolho
- chaira.*
- chalota.*
- chá-mate
- chamuscad.*

- cidrila
- coco.*-da-baía
- colza
- cominho-armênio
- conchada
- core
- coriandro
- cravinho.*
- cutelo.*
- destrinch.*
- destrinche
- echalota.*
- empelotad.*
- encaroçad.*
- endrão
- erva-cidreira-verdadeira
- erva-mate
- escalfad.*
- feij.*-rosinha.*
- feno-grego
- filé-mignon
- gaze
- gelatinos.*
- gengibre-de-dourar
- gengibre-dourado
- inoxidável
- junípero
- levístico
- ligústica
- luzendro
- macerad.*
- macis
- manjerição-cheiroso
- manjerição-dos-jardins
- manjerona-selvagem
- maracujá.*-doce.*
- melissa
- melissa-romana

- moendo
- nervo.*
- paninho.*
- péle.*
- pepininho.*
- pimenta-branca
- pimenta-preta
- pimenta-verde
- pimenta-vermelha
- pimpinela
- pincéis
- pincelad.*
- pincelada.*
- pipoca.*
- pique.*nique.*
- pires
- pistácio.*
- pleurotus
- poejo
- poejo-das-hortas
- poejo-real
- quioiô
- redução
- reduzid.*
- repolho.*-roxo.*
- rosma.*inho
- sachezinho.*
- salsa-crespa
- segurelha-de-verão
- selad.*
- sésam.*
- sobre-coxa.*
- sorgo
- suar
- talhad.*
- termômetro.*
- tocari
- trinch.*

Quando observei as UTEs com apenas um elemento que não contém palavra-chave, constatei que, curiosamente, quase todas as ervas e temperos que constam dessa lista são provenientes do glossário que produzi, juntamente com outros alunos, em 2001, como trabalho de conclusão de curso para a disciplina de Terminologia do

curso de Especialização e Tradução²¹⁴, e que haviam sido incluídas no glossário da memória de tradução.

Observei também que muitas das unidades apontadas como não sendo palavra-chave no *corpus* são, na verdade, formas concorrentes. Alguns exemplos: *badiana*, *badiana-da-china*, *badiana-de-cheiro* e *cana-do-brejo* são parassinônimos de *anis-estrelado* (que é palavra-chave); *aneto*, *endrão* e *luzendro* são parassinônimos de *endro* (palavra-chave); *coriandro* e *quioiô* são parassinônimos de coentro (palavra-chave). Todas essas variantes, aparentemente, caíram em desuso – mas isso vai ser confirmado na etapa final deste estudo, quando da contagem de sua frequência no *corpus*. Já no caso de *chalota* e *echalota*, são grafias aportuguesadas de *échalote* (do francês), assim como *wók* (do chinês transliterado); tanto *échalote* quanto *wok* são palavra-chave, embora não constem da lista de UTEs. Há também formas que, sem o hífen, contém palavra-chave: como *cebolinha francesa*, *alface lisa* e *pimenta vermelha*, *arroz selvagem*, entre outras, o que já é um indício da preferência de uso das formas não-hifenizadas nas receitas do *corpus*.

Por outro lado, chama a atenção o fato de algumas formas concorrentes terem sido apontadas ambas como palavra-chave: *agar agar* / *agar-agar*, *maizena* / *maisena* / *amido de milho*, *alho-poró* / *alho porro*, entre outras. Algumas UTEs, por outro lado, não contém, de fato, nenhuma palavra-chave, ainda que as duas grafias possíveis tenham sido testadas – como é o caso de *feno-grego* e *feno grego*, ou dos radicais “chamusc-”, “pincelad-” e “destrinch-”. Resta saber sua frequência no *corpus* para que se possa afirmar se se trata de lexias já em desuso nesse tipo de texto.

Por fim, identifiquei com pesar a presença de *enquanto isso* nessa lista de UTEs que não contém palavras-chave. Digo “com pesar” não por acreditar que um ou ambos elementos deveriam ser palavra-chave – pela alta frequência que têm na língua geral, tais palavras dificilmente alcançariam, individualmente, esse status em textos culinários²¹⁵ – mas por uma questão metodológica. Se à primeira vista parece óbvio

²¹⁴ disponível em: http://www.fflch.usp.br/citrat/glossario/gloss/arquivo_glossario11b.htm [acesso em: 15/7/08]

²¹⁵ fiz um teste usando *wordlists* de 2 palavras do *corpus* de estudo e de referência para elaborar uma lista de palavras-chave: “enquanto isso” é palavra-chave dupla e ocupa a posição 658 da lista. Isto é, em vez de concentrarmos nas limitações impostas pelo uso da lista de palavras-chave como ponto de partida para a busca de UTEs no *corpus* – tais como a desconsideração de palavras de baixa frequência ou, pelo contrário, de palavras de frequência muito alta na língua geral – poderíamos ampliar seu uso para englobar, por exemplo, listas de palavras-chave de 2 ou até mesmo de 3 elementos.

concluir que um dos quesitos importantes e confiáveis a ser empregado na identificação de novas UTEs no *corpus* seria verificar se a palavra extraída é ou não uma palavra-chave, uma UTE como “enquanto isso” me faz querer repensar a confiabilidade de se usar esse critério, sozinho, na extração de novos candidatos a UTEs.

Embora “enquanto” e “isso” não sejam palavras de conteúdo, juntas, constituem um padrão importante para a produção de textos da área – prova disso é o fato de, em português europeu, a expressão usada na mesma situação ser “entretanto” (Tagnin e Teixeira, 2004). Isso parece apontar, de fato, para a necessidade de se combinar vários critérios como os levantados no presente estudo a fim de se obter um método eficiente de extração semi-automática de UTEs do *corpus* já que, pelo menos no caso de “enquanto isso”, é a ocorrência **conjunta** dessas duas palavras que pode ser considerada uma Unidade de Tradução Especializada, e não sua ocorrência isolada. De qualquer modo, acredito que, ainda que se considerasse apenas a lista de palavras-chave com 1 unidade, estaríamos ganhando mais que perdendo na extração de UTEs, ao menos do ponto de vista da tradução.

4.2.4 Frequência das UTEs no *corpus* de português

Um fator importante a ser considerado em estudos baseados em *corpora* é a frequência (Berber Sardinha, 2004: 162). Como as UTEs provenientes de minha memória de tradução foram compiladas sem nenhum critério quantitativo, julguei importante verificar a frequência de cada uma delas no *corpus* de Culinária, para que a probabilidade de ocorrência pudesse ser levada em conta e, portanto, a relevância dos dados levantados a partir do estudo dessas UTEs.

4.2.4.1 Verificando a frequência das UTEs no *corpus* total

Para esta etapa do estudo também utilizei uma rotina em Shell (caso contrário, teria que procurar cada uma das 1.294 UTEs no corpus, uma a uma, usando uma ferramenta como o *Concord* do WordSmith Tools). Substituí as porções variáveis das palavras por .* (ponto asterisco) com o intuito de indicar para o programa que qualquer sequência de caracteres poderia ocupar aquela porção da expressão de busca. A aparência do arquivo, nomeado *UTES_todas.txt*, pode ser vista na figura 62.

Figura 62: UTEs com .* substituindo a parte variável

Outra etapa necessária compreendeu a reunião de todos as receitas de todos os *subcorpora* num único arquivo (*corpTese.PB.TT.txt*) – que foi bastante trabalhosa, pois os mesmos continham inúmeras subpastas e, muitas vezes, tive que fazer correções em lote nos arquivos originais antes de colá-los no arquivo total, eliminando etiquetas HTML que não tinham sido apagadas pelos métodos semi-automáticos de limpeza de *corpus* usados²¹⁶. O arquivo resultante traz todas as receitas do *corpus* coladas uma após a outra.

Isso feito, o passo seguinte foi escrever a rotina e rodá-la no Cygwin. Uma rotina semelhante à usada para encontrar as palavras-chave entre as UTEs foi usada:

```
#!/bin/sh
while read UTEs
do
    freq=$(grep -c -i "$UTES" corpTese.PB.TT.txt)
    echo -e $UTES '\t' $freq
done < UTEs_todas.txt > freq2.UTES.TT.txt
```

O arquivo resultante (*freq2.UTES.TT.txt*) mostra uma UTE por linha, na coluna da esquerda, e a frequência da mesma na coluna da direita. Para colar o resultado na tabela Excel, copiei todo o resultado para um arquivo do Word, solicitei a transformação

²¹⁶ para maiores informações sobre limpeza em lote de arquivos para compilação de corpus, vide Berber Sardinha, 2004 (capítulo 2).

do texto em tabela (usando como separador a marca de tabulação), apaguei a coluna da esquerda (com as UTEs), copiei a coluna restante (com a frequência das UTEs) e coleí na coluna reservada²¹⁴. A aparência final da tabela pode ser vista na figura 63.

unidade de tradução terminológica						Cat. Gram. Itens	Cat. Gram. Geral	no. de elemento	contém pal. Chave	freq. no corpus
1	2	3	4	5	6	7	8	9	10	11
a gosto	à Juliana	abacaxi*	abaf*	abaix* a massa	abaix* o fogo	abobrinha* italiana*	abr* a massa	abr* a massa com o rolo	abr* a massa sobre uma superfície enfarinhada	abr* a massa sobre uma superfície polvilhada com farinha
prp + NOM	PRP&DET + NOM	NOM	v	V + DET + NOM	V + DET + NOM	NOM + ADJ	V + DET + NOM	V + DET + NOM + PRP + DET + NOM	V + DET + NOM + PRP + DET + NOM + ADJ	V + DET + NOM + PRP + DET + NOM + ADJ + PRP + NOM
SAdv	SAdv	subs	SV	SV	SV	SN	SV	SV	SV	SV
2	2	1	1	3	3	2	3	6	7	9
sim	não	sim	sim	sim	sim	sim	sim	sim	sim	sim
4425	3	538	32	26	210	29	444	15	2	4

Figura 63: aparência da tabela final

4.2.2.5 Análise dos dados e algumas considerações preliminares

A tabela a seguir mostra a frequência geral das UTEs no *corpus* de Culinária português. Por exemplo: apenas 3 UTEs da lista ocorrem mais de 10.000 vezes no corpus, ao passo que a grande maioria (290 UTEs) ocorreram entre 11 e 50 vezes.

no. de ocorrências corpus	no. de UTEs na faixa
acima de 10.001	3
entre 5.001 e 10.000	12
entre 2.501 e 5.000	23
entre 1.001 e 2.500	52
entre 501 e 1.000	76
entre 251 e 500	94
entre 101 e 250	160
entre 51 e 100	121
entre 11 e 50	290
entre 6 e 10	50

²¹⁴ vide item 4.2.2 para maiores detalhes sobre esse procedimento.

entre 3 e 5	102
2	47
1	70
0	137

Tabela 15: faixas de freqüência das UTEs

Como eu esperava, 55 das 137 UTEs com 0 oc. no *corpus* (40%), estão na lista de UTEs que não contêm palavra-chave (vide abaixo) e, com exceção de três (aos *bocados*, *feno grego* e *balança digital*), são todas monovocabulares. Isso evidencia, por um lado, a relação entre associação de palavras e importância (chavicidade) para o léxico de uma área. De outro, ressalta a importância de se considerar a freqüência de uso para a caracterização dos padrões léxico-gramaticais que tipificam gêneros e tipologias textuais; isto é, itens lexicais que têm ocorrência muito baixa ou igual a zero num *corpus* representativo, sejam termos ou não, terão pouca importância, do ponto de vista da probabilidade, para a caracterização dos padrões típicos da área (sua inclusão em dicionários voltados para o produtor de textos, nesse sentido, seria questionável).

À primeira vista, somente as palavras em negrito nesta lista seriam problemáticas, caso se optasse por adotar como único critério a eliminação de quaisquer palavras do *corpus* com ocorrência igual a 0 que não fossem chave, pois são termos raros, mas não em desuso na culinária. Seu status e freqüência mudaria, é possível, se o tamanho do *corpus* fosse aumentado consideravelmente.

alecrim-de-cheiro	coriandro	→ pepininho.*
alfavaca-do-campo	cutelo.*	→ pimenta-preta
anis-doce	endrão	pimpinela
aos bocados	erva-cidreira-verdadeira	→ pincéis
badiana	escalfad*	pistácio*
badiana-da-china	→ feno grego	poejo-das-hortas
badiana-de-cheiro	→ feno-grego	poejo-real
→ balança digital	gengibre-de-dourar	quioiô
→ banana*-passa*	gengibre-dourado	sachezinho.*
batata.*-rosada.*	→ junípero	segurelha-de-verão
cana-do-brejo	levístico	selad.*
capim-de-cheiro	ligústica	sésam.*
cariz	luzendro	sorgo
cerefolho	→ macis	suar
chaira.*	manjerição-cheiroso	tocari
cidrila	manjerição-dos-jardins	→ wók
coco.*-da-baía	manjerona-selvagem	
colza	melissa-romana	
cominho-armênio	péle.*	
conchada		

Uma consulta manual ao *corpus* usando o *Concordance*, no entanto, mostra que algumas delas poderiam ser eliminadas dessa lista: há 12 ocorrências de “banana passa”, sem o hífen, no *corpus* (1 no plural), 2 de “fenogrego”, 22 de “pimenta preta”, “pincel”, no singular, ocorre 19 vezes, “cerefólio” (não “cerefólio”) ocorre 12 vezes, e “wok” sem acento 51 vezes. Ou seja, as ocorrências não foram encontradas por uma questão metodológica relacionada à busca (que não previu variações ortográficas), e não porque não ocorrem no *corpus*.

A consulta manual foi usada também para checar algumas frequências obtidas que causaram estranheza: “abr.* a massa sobre uma superfície enfarinhada” teria apenas 2 ocorrências no *corpus*? Uma busca por “abr*” no programa Concordance com “enfarinhada” ocorrendo até a 7a. palavra à direita retornou o seguinte resultado:

N	Concordance
3	a geladeira por 30 minutos. Em seguida, divida a massa em 2 partes e abra em uma superfície enfarinhada até formar dois retângulos de 40 x
4	s. Ligue o forno à temperatura média. Divida a massa em 26 porções, abra-as numa superfície enfarinhada com as mãos, formando um círculo
5	rve. Ligue o forno à temperatura média. Retire a massa da geladeira e abra-a numa superfície lisa e enfarinhada até obter um retângulo de 25
6	dos ingredientes sendo a farinha de trigo aos poucos. Amasse bem e abra a massa com rolo numa mesa enfarinhada formando um retângulo
7	e leve à geladeira por 12 horas. Retire metade da massa da geladeira, abra com um rolo sobre uma superfície enfarinhada até obter a espess
8	e tomate, o sal, o orégano e a pimenta e refogue. Retire e deixe esfriar. Abra a massa em superfície enfarinhada e recorte-a com uma xícara g
9	lisa e sove até desgrudar das mãos. Deixe descansar por 15 minutos. Abra a massa numa superfície enfarinhada até obter 1 cm de altura. C
10	nsar em lugar quente, por cerca de 20 minutos, até dobrar de tamanho. Abrir a massa numa superfície enfarinhada e forrar uma fôrma redonda
11	nsar em lugar quente, por cerca de 20 minutos, até dobrar de tamanho. Abrir a massa numa superfície enfarinhada e forrar uma fôrma redonda
12	e tomate, o sal, o orégano e a pimenta e refogue. Retire e deixe esfriar. Abra a massa em superfície enfarinhada e recorte-a com uma xícara g
13	etire do fogo e reserve. Ligue o forno à temperatura média. Montagem: abra a massa em uma superfície enfarinhada no formato de um retângulo
14	ebola e o restante dos ingredientes. Esta mistura deve ficar bem seca. Abra a massa numa superfície lisa enfarinhada (ou na máquina) e corte
15	e leve à geladeira por 30 minutos. Ligue o forno à temperatura média. Abra a massa numa superfície lisa e enfarinhada numa espessura de 0
16	ínio, de 10 a 15 cm de diâmetro, próprias para empadões. Montagem: abra a massa em uma superfície enfarinhada e use ¾ dela para forrar
17	Por último, misture o zattar e retire. Ligue o forno à temperatura média. Abra a massa em uma superfície enfarinhada e corte-a em 20 quadrad
18	, ou até secar. Acerte o sal, acrescente a salsinha, retire e deixe esfriar. Abra a massa em uma superfície enfarinhada e corte 9 círculos de 10 c
19	re a massa. Embrulhe em filme plástico e leve à geladeira por 2 horas. Abra a massa em uma superfície levemente enfarinhada e com ela forr
20	orme as instruções da embalagem. Ligue o forno à temperatura média. Abra a massa sobre uma superfície enfarinhada e forre, em seguida, u
21	lher (sopa) de azeite de oliva e o sal. Misture vigorosamente e reserve. Abra a massa em uma superfície enfarinhada com um rolo ou um cilind
22	ture bem e retire do fogo. Reserve. Ligue o forno à temperatura média. Abra a massa em uma superfície enfarinhada e, com o auxílio de um c
23	à geladeira por 30 minutos. Ao final do tempo, com o auxílio de um rolo, abra a massa sobre uma superfície ligeiramente enfarinhada e forre u
24	à geladeira por 30 minutos. Ao final do tempo, com o auxílio de um rolo, abra a massa sobre uma superfície ligeiramente enfarinhada e forre u
25	fogo e misture o peixe e o camarão. Acerte o sal e reserve. Montagem: abra a massa em uma superfície enfarinhada em 6 círculos de 16 cm d
26	e crescer, em local aquecido, por 30 minutos, ou até dobrar de volume. Abra a massa grosseiramente em uma superfície enfarinhada, espalhe
27	do água até formar uma pasta. Reserve. Divida a massa em 2 partes e abra-a em uma mesa enfarinhada, em formato retangular. Espalhe o re
28	amornar. Bata em um processador a ricota, a mussarela e o requeijão. Abra a massa em uma mesa enfarinhada, em espessura fina. Com a
29	cada um em 12 fatias. 3. Faça uma bola com cada fatia. Com um rolo, abra cada bola sobre uma superfície enfarinhada, formando discos. Es
30	, até encorpar e perder o líquido. Retire do fogo, acerte o sal e reserve. Abra a massa em uma superfície enfarinhada, corte 10 quadrados de
31	deixe refogar por mais 3 minutos. Acerte o sal, retire do fogo e reserve. Abra a massa sobre uma superfície enfarinhada, de modo a ficar bem f
32	em fogo baixo por 20 minutos, ou até o molho encorpar. Enquanto isso, abra a massa fina em superfície enfarinhada. Corte 17 retângulos (12 x
33	plástico e leve à geladeira por 30 minutos. Com um rolo de macarrão, abra a massa sobre uma superfície ligeiramente enfarinhada. Forre um

Figura 64: linhas de concordância para “abr*” + “enfarinhada”

Como é possível observar, o programa retornou o número correto de ocorrências da expressão *abr* a massa sobre uma superfície enfarinhada* (linhas 20 e 31), mas a busca não prevê uma série de UTEs que apresentam ligeiras alterações do padrão “abrir + art + MASSA + prep + SUPERFÍCIE + enfarinhada” – que só são recuperadas numa busca manual. Esse é o tipo de estrutura de que as metodologias totalmente automáticas costumam não dar conta. O ideal, nesse caso, seria pesquisar o *corpus*

usando expressões regulares que previssem todas essas variações (ou a maioria delas). Mas julgo que seria uma metodologia muito trabalhosa apenas para cumprir esta etapa do estudo. Reservarei o uso de expressões regulares para buscar padrões morfossintáticos no *corpus* – o que retornaria muito mais resultados do que uma expressão regular dedicada apenas a encontrar as variações de uma única UTE (que podem ser facilmente recuperadas com uma busca manual no aplicativo Concordance do *WordSmith Tools*).

Retomemos a análise das freqüências: há 14 UTEs monovocabulares que não contêm nenhuma palavra-chave mas ocorrem 1 vez no *corpus* (*arroz-selvagem*, *batata.*-roxa.**, *chamusc.**, *chamuscad.**, *destrinch.**, *encaroçad.**, *erva-mate*, *gelatinos.**, *moendo*, *paninho.**, *poejo*, *repolho.*-roxo.**, *rosma.*inho*, *salsa-crespa*) – para algumas delas, como *salsa-crespa*, isso se deveu à grafia hifenizada, já que *salsa* é palavra-chave. Em situação semelhante, mas ocorrendo 2 vezes no *corpus*, há: *almofariz.**, *batata.*-bolinha*, *cebolinha-verde*, *echalota.**, *empelotad.**, *maracujá.*-doce.**. Ainda relacionando freqüência e palavra-chave, observou-se que entre as 113 UTEs que não contêm nenhuma palavra-chave, há apenas 5 com mais de 10 ocorrências no *corpus*: *enquanto isso* (321 oc.), *aos pouquinhos* (42 oc.), *reduzid.** (30 oc.), *pipoca.** (14 oc.) e *pires* (11 oc.). Isso, mais uma vez, atesta a importância de se usar critérios combinados na extração semi-automática de UTEs, caso contrário, uma UTE como “enquanto isso”, com uma freqüência tão significativa no *corpus*, seria perdida.

Se, combinando todos os critérios empregados na confecção da tabela, selecionássemos apenas: i) freqüência no *corpus*: igual ou maior que 2; ii) contém palavra-chave?: sim; iii) número de elementos: 2 ou mais, restariam 685 UTEs da lista original de cerca de 1.300. Em termos de padrões morfossintáticos, teríamos 107 distintos (dos 135 existentes na lista com todas as UTEs). Resta saber se a freqüência é um critério adequado para eliminar padrões morfossintáticos, já que a chavidade é um critério que se refere não ao padrão, mas às palavras individuais que o compõem, e portanto não se aplica nesse caso.

Tomemos como exemplo a UTE “até que evapore”, que teve 0 oc. no *corpus*. Uma busca usando o utilitário Concord revela que há variações dessa UTE no *corpus*:

N	Concordance
1	enta-do-reino e acrescente 1 xícara de água. Cozinhe em fogo baixo até que a água evapore. Reserve. Prepare o molho pico de galo, ba
2	Regue com o vinho e retorne todos os escalopes à frigideira, cozinhe até que metade do vinho evapore, regue com o caldo e cozinhe até q
3	ma panela com o vinho, o vinagre e a pimenta. Reduza em fogo baixo até que o líquido evapore quase que completamente (cerca de 1 colh
4	que o vinho, o tomilho e uma pitada de sal. Reduza o fogo e cozinhe até que o líquido evapore totalmente. Deixe esfriar. Coloque o queij
5	fogo baixo. Lavar o espinafre e juntar com alho, cozinhar suavemente até que o líquido evapore. Juntar o espinafre com o risoto de quinoa.
6	inutos. Adicione o caldo e os ramos de tomilho e cozinhe em fogo alto até que o líquido se evapore por completo. Acrescente o creme de lei
7	m uma colher de pau, dissolvendo todos os resíduos. Deixe cozinhar até que o líquido se evapore quase completamente. Junte o creme de
8	çar a dourar, regue-a aos poucos com a vinha d'alhos e deixe-a assar até que o líquido se evapore totalmente. Sirva o prato ainda quente ac
9	é terminar. Despeje por cima o vinagre e o óleo. Leve ao fogo e deixe até que o vinagre evapore, ficando apenas o óleo. Estão prontas as
10	minutos. Adicione então 1 copo de vinho branco seco. Deixe cozinhar até que o vinho evapore pela metade. Numa xícara dissolva 1 colher,
11	gue por 2 minutos e acrescente o vinho. Mexa com uma colher de pau até que o vinho evapore. Junte a polpa das pétalas e o coração das a
12	fogo baixo e acrescente os tomates batidos com o pimentão, refogue até que se evapore o máximo de líquido desses ingredientes. Rale f

Figura 65: linhas de concordância para “até que... evapora”

Uma busca no *corpus* etiquetado usando uma expressão regular que permite a ocorrência de 0 a 30 caracteres entre o pronome relativo e o verbo – `\s+(\S+)/prp\s+(\S+)/pr\s+([\S+\S+\s+]{0,30}?)\s+(\S+)/v\s+`²¹⁵ – retorna 2.217 ocorrências. Agrupadas as repetições, foram 811 UTEs distintas; se forem eliminadas as que tiveram frequência abaixo de 3, restam 127 UTEs, listadas a seguir, em ordem alfabética (o número de ocorrências no *corpus* está entre parênteses, logo após cada uma delas).

²¹⁵ `\s+` = pelo menos um espaço em branco; `(\S+)` = pelo menos um caractere diferente de espaço em branco (os parênteses indicam a porção recuperada que deve ser exibida no resultado); `([\S+\S+\s+]{0,30}?)` = qualquer seqüência de letras seguida de qualquer etiqueta morfosintática iniciada por “/” seguida de um ou mais espaços em branco até o total de 30 caracteres. Mais detalhes sobre a construção de expressões regulares e seu uso na linguagem de programação Python serão dados no item 4.2.5.

até que a água seque (5)
 até que a batata esteja (4)
 até que a carne esteja (35)
 até que a carne fique (8)
 até que a cebola esteja (33)
 até que a cebola fique (3)
 até que a cebola murche (5)
 até que a couve esteja (3)
 até que a farinha esteja (3)
 até que a gelatina esteja (3)
 até que a maçã esteja (3)
 até que a massa comece (3)
 até que a massa esteja (30)
 até que a massa fique (7)
 até que a massa se desprenda (3)
 até que a massa se solte (3)
 até que a mistura engrosse (4)
 até que a mistura esteja (8)
 até que a mistura fique (9)
 até que a pele esteja (3)
 até que a superfície esteja (7)
 até que a torta esteja (3)
 até que ao enfiar (14)
 até que ao espetar (5)
 até que ao ser (19)
 até que apareça (6)
 até que as batatas estejam (6)
 até que as cebolas estejam (3)
 até que as fatias estejam (8)
 até que comece (23)
 até que comecem (7)
até que derreta (6)
 até que desgrude (3)
 até que dobre (16)
 até que dobrem (4)
até que doure (8)
até que dourem (13)
até que endureça (7)
 até que enfiando (117)
até que engrosse (10)
até que esfrie (3)
 até que esteja (163)

até que estejam (140)
até que ferva (9)
 até que fique (87)
 até que fiquem (43)
 até que forme (4)
até que murche (6)
até que murchem (4)
 até que o açúcar derreta (7)
 até que o açúcar dissolva (10)
 até que o açúcar esteja (3)
 até que o arroz esteja (17)
 até que o aspargo esteja (3)
 até que o bolo esteja (8)
 até que o creme engrosse (6)
 até que o espinafre esteja (3)
 até que o feijão esteja (4)
 até que o frango esteja (6)
 até que o líquido seque (4)
 até que o líquido tenha (4)
 até que o molho engrosse (7)
 até que o molho esteja (9)
 até que o molho se reduza (3)
 até que o peixe esteja (9)
 até que o queijo derreta (3)
 até que o queijo esteja (6)
 até que o recheio esteja (5)
 até que o salmão esteja (3)
 até que o tomate esteja (5)
 até que os biscoitos estejam (3)
 até que os bolinhos estejam (4)
 até que os filés percam (3)
 até que os grãos estejam (13)
 até que os ingredientes estejam (13)
 até que os legumes estejam (10)
 até que os pimentões estejam (3)
 até que os tomates comecem (3)
 até que os vegetais estejam (6)
 até que se forme (6)
 até que se incorporem (4)
 até que solte (4)
 até que terminem (6)
 deles e cubra (3)

do que sagrado (3)
em que cozinhou (5)
em que dourou (3)
 em que ficaram (4)
 em que ficou (9)
 em que foi (5)
 em que for (5)
em que fritou (6)
 em que irá (14)
 em que os funghi ficaram (4)
 em que será (5)
 em que vai (4)
 para que a cebola fique (20)
 para que a farinha doure (3)
 para que a manteiga derreta (3)
 para que a massa fique (6)
 para que a mistura fique (6)
para que asse (4)
para que assem (7)
para que dourem (19)
 para que ela fique (3)
para que engrosse (4)
 para que fique (42)
 para que fiquem (34)
 para que não escureçam (6)
 para que não fiquem (4)
 para que não grude (4)
 para que não grudem (4)
 para que o açúcar desmanche (3)
 para que o açúcar dissolva (9)
 para que o arroz fique (3)
 para que o molho engrosse (3)
 para que o pão absorva (3)
 para que o queijo derreta (7)
 para que o recheio fique (3)
 para que o recheio não escape (3)
 para que os ingredientes comecem (3)
 para que os ovos cozinhem (4)
para que sequem (3)
 para se preparar (5)
sem que ferva (3)
 sem se importar (4)

Uma leitura ainda que superficial dessa lista permite a observação vários dados importantes:

- embora “até que evapore” não ocorra no corpus, o padrão “prp + pr + v” é bastante produtivo – outras UTEs com esse mesmo padrão (vide formas em negrito na lista) totalizam 114 ocorrências (fora as que tem ocorrência 1 e 2, descartadas da lista);

- o padrão “prp + pr + v” admite várias estruturas entre a locução prepositiva “prp + pr” e o verbo = (prp + pr (+ ...) + v)
- os verbos e sintagmas verbais que ocorrem na estrutura “até que [+ART(def.) + N] + V” podem:

→ **descrever um processo** – *dourar, fritar, dissolver, derreter, dourar por igual*, etc., no pres. do subjuntivo;

→ **ser verbos de suporte** – *fique(m), esteja(m), comece(m) a (ficar)* – que introduzem um estado, indicado por um adjetivo ou sintagma adjetival (*macio, ligeiramente dourado, lisa e elástica, homogêneo*, etc.).

- Uma consulta ao *corpus* usando a ferramenta *Concord* revela que a UTE “até que [+ART(def.) + N] + V”, em geral, é precedida de uma que pode ser descrita, em termos gerais, por “V(processo-imperativo) por SN(TEMPO)”:

N	Concordance
158	ueijo. vDistribua entre as forminhas e asse os suflês por uns 15 minutos, até que estejam crescidos e bem dourados na superfície e nas laterais.
159	s ervas amarrados com um barbante. * Asse por mais ou menos 1 hora, até que o tomate esteja macio e perfumado, deixe esfriar e guarde na ge
160	, junte as pontas do papel, torça para fechar e asse por cerca de 1 hora, até que a polpa esteja bem macia e um perfume delicioso invada a cozin
161	tar uma na outra. * Asse a primeira leva de grissinis por uns 20 minutos, até que estejam bem dourados e crocantes, enquanto prepara o restante
162	imento com a massa restante. * Asse as trouxinhas por uns 10 minutos, até que estejam douradas e crocantes, passe para um prato e regue co
163	do. * Cozinhe a massa em água fervente e salgada por uns 10 minutos, até que esteja “al dente” (cozida, mas resistente à mordida). * Escorra,
164	com sal e pimenta, regue com azeite e leve ao forno por uns 30 minutos, até que esteja macio e ligeiramente dourado. * Cozinhe a massa em ág
165	te, o louro e 1 pitada de sal e cozinhe em fogo baixo por uns 30 minutos, até que o tomate esteja macio e saboroso, então acerte o sal, descarte
166	de sopa de sal e a fava assim que ferver e cozinhe por uns 10 minutos, até que esteja macia, mas ainda bem verde. * Escorra a água do cozim
167	baixe a temperatura para 200°C (médio-alto) e asse por uns 25 minutos, até que os pães estejam crescidos e dourados. * Retire do forno, esper
168	e então o sal e o azeite e vá juntando aos poucos a farinha e trabalhando até que a massa fique bem macia e se solte das mãos. * Cubra com u
169	nta numa assadeira média, espalhe e leve ao forno por 10 a 15 minutos, até que esteja dourado e crocante. * Junte as raspas de limão, misture
170	no por mais ou menos 1 hora e 15 minutos, mexendo de vez em quando, até que os pedaços estejam bem dourados e macios. * Descarte as er
171	umente a temperatura para 220°C (alto) e asse por mais uns 20 minutos, até que o frango esteja bem dourado e macio (espete a coxa com um ga
172	ma panela e deixe ferver. Mergulhe os tomates aos poucos e escale-os até que a pele solte. Escorra e espere esfriar para retirar a pele e as se
173	uira uma textura de areia grossa, acrescente a água gelada aos poucos até que fique quase ligada. Retire, aperte bem, coloque em um saco plá
174	dor, coloque a farinha de trigo, sal, a manteiga e a gordura vegetal. Pulse até que a farinha adquira uma textura de areia grossa, acrescente a água
175	finas e reserve. MOLHO Misture todos os ingredientes e leve ao fogo até que ferver. Coloque as peras fatiadas nesse líquido e deixe por algu
176	Espalhe as amêndoas sobre as tortas e leve ao forno por 30 minutos ou até que a massa fique dourada e a mistura de geléia borbulhe. Retire do
177	cer em fogo médio, até atingir o ponto de fervura. Ferva por 2 minutos ou até que a mistura engrosse levemente, mexendo sempre. Derrame o mol
178	rescente mais manteiga e frite-os (poucos por vez) durante 3 minutos, ou até que os filês percam a cor rosada. Vire-os uma vez e coloque em um
179	do peito na horizontal, em dois pedaços. Bata-os com martelo de carne até que cada um atinja a espessura de 0,25cm. Polvilhe-os com sal. Em

Figura 66: linhas de concordância para “até que”

- há casos de uso de pronome reflexivo: “até que se incorporem”, “até que se forme...”, “até que a massa se desprendam”;
- algumas estruturas destoam das demais, como: “até que **ao** enfiar”, “até que **ao** espetar”, “até que **ao** ser”, “até que **enfiando**” – todas elas representativas de uma fraseologia muito recorrente no corpus, e com alto grau de variação estrutural, que descreve a técnica de espetar um alimento com objeto pontiagudo / cortante para verificar o ponto de cozimento. É usada para testar bolos, tortas e/ou pudins, carnes e peixes assados (no primeiro caso, o objeto

deve sair seco / limpo, no segundo, o suco da carne deve sair límpido e claro, sem sinais de sangue; se for peixe, a carne deve separar-se em lascas). Eis alguns exemplos retirados do *corpus* para “até que... ao”:

N	Concordance
6	gando de vez em quando com o molho da assadeira, por 30 minutos ou até que, ao ser espetada com um garfo, a carne se separe em lascas.
7	gano restante, a casca de limão e sal. 5. Asse por mais 15 minutos ou até que, ao ser espetada com um garfo, a carne se parta em lascas. 4
8	xes. Cubra-os com os pedaços de manteiga. 5. Asse por 25 minutos ou até que, ao ser espetada com um garfo, a carne do peixe se solte em la
9	o baixo, disponha os filés e frite, virando-os às vezes, por 15 minutos ou até que, ao ser espetada com um garfo, a carne se separe em lascas.
10	e o peixe, virando-o de vez em quando, por 15 minutos de cada lado ou até que, ao ser espetada com um garfo, a carne se separe em lascas.
11	macia. Junte o peixe, o suco de limão e o sal. Cozinhe por 5 minutos ou até que, ao ser espetada com um garfo, a carne se desfaça em lascas.
12	al ao fogo alto e deixe ferver. Junte o peixe e cozinhe por 10 minutos ou até que, ao ser espetada com um garfo, a carne se parta facilmente. Es
13	pel impermeável untado. Leve ao forno pré-aquecido por 20 minutos ou até que, ao se enfiar um palito, este saia seco. Enquanto isso, misture o
14	e o leite. Coloque a mistura na fôrma preparada e asse em banho-maria até que, ao inserir uma faca, ela saia limpa (cerca de 40 minutos). Dei
15	ar em forno preaquecido a 170° C por aproximadamente 45 minutos, ou até que, ao inserir um palito no centro do bolo, ele saia limpo. Retire do
16	com quadradinhos irregulares). Leve para assar em forno preaquecido, até que, ao inserir um palito no centro da massa, este saia limpo, indic
17	ma assadeira untada e polvilhada com farinha de trigo. Leve para assar até que, ao inserir um palito no centro da massa, este saia limpo. Caso
18	orminhas caramelizadas com açúcar. Leve para assar em banho-maria até que, ao inserir um palito no centro do pudim, este saia limpo. Leve
19	dientes restantes. Distribua sobre o badejo. 4. Asse por 30 minutos ou até que, ao espetar uma ponta do peixe com um garfo, ela se separe e
20	Asse o peixe, coberto com o papel-alumínio, por cerca de 30 minutos ou até que, ao espetar uma ponta com um garfo, ela se separe em lascas.
21	senforme. 6. Regue a tainha com a manteiga, a água e o vinho e asse até que, ao espetar uma ponta com um garfo, ela se separe em lascas.
22	io. 4. Asse, regando com o caldo de vez em quando, por 40 minutos ou até que, ao espetar com um garfo, a carne se separe em lascas. Aco
23	o forno em banho-maria a uma temperatura alta, por uns 45 minutos, ou até que, ao espetar com um garfo, ele saia limpo. Sirva morno ou frio.
24	alto. Junte as trutas e polvilhe com sal. Tampe e cozinhe em fogo baixo até que, ao espetar a carne do peixe com um garfo, ela se solte com fa
25	a forma. Cozinhe no microondas, em potência média, por 15 minutos ou até que, ao enfiar um palito, ele saia seco. Deixe descansar por 10 min
26	. Coloque na forma e asse em banho-maria por cerca de 45 minutos ou até que, ao enfiar um palito, ele saia seco. Desenforme. Molho 1. En
27	. Adicione o coco e misture. Coloque na forma e asse em banho-maria até que, ao enfiar um palito, ele saia seco (cerca de 40 minutos). Deixe
28	loque-as numa assadeira. Asse no forno preaquecido, em banho-maria, até que, ao enfiar um palito na massa, ele saia seco. Deixe esfriar e de
29	ientes do pudim e coe. 2. Distribua nas formas e asse em banho-maria até que, ao enfiar um palito próximo à borda, ele saia seco (cerca de 35
30	tando fundo e lados. Coloque a massa por cima e asse em banho-maria até que, ao enfiar um palito próximo à borda, ele saia seco (cerca de 30
31	eservada e leve ao forno, em banho-maria. Asse em temperatura média até que, ao enfiar um palito próximo à borda, ele saia seco (cerca de 45
32	ure bem. Distribua entre as formas reservadas e asse, em banho-maria, até que, ao enfiar um palito próximo à borda, ele saia seco (cerca de 35
33	ata até ficar homogêneo. 5. Coloque na forma e asse em banho-maria até que, ao enfiar um palito próximo à borda, ele saia seco (cerca de 1
34	garfo em ziguezague na superfície para marmorizar o creme. 4. Asse até que, ao enfiar um palito no centro da torta, ele saia seco (cerca de 1
35	te a baunilha e acrescente os ovos, um a um. 2. Recheie a torta e asse até que, ao enfiar um palito no centro, ele saia seco (cerca de 1 hora).
36	car homogêneo. 3. Coloque a massa na forma. Asse por 40 minutos ou até que, ao enfiar um palito no centro, ele saia seco. Deixe amornar.

Figura 67: linhas de concordância para “até que... ao”

- a associação “prp + pr” não se aplica somente a “até que”, mas também a “para que”, “em que”, “sem que” – todas, a princípio, também indicativas de UTEs da área. Por exemplo:

→ **para que [+ ART(def.) + N] [+ não] + V(processo – pres. subjuntivo)**

Ex: “para que asse”, “para que os ovos cozinchem”, “para que o açúcar dissolva”, “para que o recheio não escape”, “para que não grudem”

→ **para que [+ART(def.) + N] [+ não] + FICAR + [ADJ / LOC. ADJ.]**

Ex: “para que o arroz fique...” (*solto, bem cremoso*, etc.), “para que não fiquem...” (*moles demais, amargas*, etc.)

Dessa breve análise do padrão morfossintático de “até que evapore”, UTE que não teve nenhuma ocorrência no *corpus* de estudo, fica claro, por um lado, que a riqueza de dados obtidos por meio da busca de outras UTEs no *corpus* utilizando o

padrão “prp + pr + v” justifica plenamente ignorar-se o critério da frequência no *corpus* como justificativa para eliminar padrões morfossintáticos de busca. Assim, serão buscadas no corpus, a princípio, ocorrências para todos os 135 padrões morfossintáticos identificados nesta etapa do estudo.

Por outro lado, essa pequena amostra de dados, a exemplo do que foi observado no estudo do termo “flour” no *corpus* de língua inglesa, dá uma idéia da complexidade envolvida no levantamento e tratamento dos dados resultantes de uma extração como essa – etapa da metodologia de extração de UTEs, que descrevo no subitem a seguir.

4.2.5 Frequência dos padrões morfossintáticos no *corpus* total

Antes de dar início ao levantamento da frequência dos padrões morfossintáticos identificados na segunda fase do estudo no *corpus* de Culinária português, procurei me familiarizar com a linguagem de programação *Python* que, assim como o *Perl* e o *Rubby*, é uma ferramenta potente de manipulação de dados lingüísticos usada por muitos lingüistas de corpus, lingüistas computacionais e programadores que trabalham com recuperação de dados textuais.

O *Python* é uma linguagem de programação de alto nível²¹⁶ e orientada a objetos, isto é, possui funcionalidades inerentes a certos tipos de dados (objetos) que facilitam a pesquisa textual. É como se cada tipo de dado, a exemplo de um aparelho eletrônico (como uma máquina de lavar roupa), viesse com botões específicos que, bastando apertá-los, dão acesso, ou desencadeiam uma série de operações envolvidas em uma função complexa (centrifugar, por exemplo). Assim, uma vez classificados, os dados textuais podem ser facilmente manipulados – uma lista de palavras, por exemplo, em que a cada item *x* corresponda um item *y* (chamada de “dicionário” em *Python*, mas que pode servir para armazenar quaisquer dados que mantenham essa relação $x \rightarrow y$, como uma palavra e sua frequência no corpus, ou sua categoria gramatical) tem funcionalidades (“botões”) que permitem, com um simples comando, fazer buscas complexas, reordenações, adição e eliminação de itens ou conjuntos de itens, etc.

²¹⁶ a sintaxe dos comandos se assemelha mais à linguagem humana que à linguagem de máquina (0 e 1) – são portanto mais compreensíveis.

Como a maioria dos programas escritos para ambiente Unix / Linux, a linguagem *Python* é um software livre – qualquer pessoa pode baixar o código da Internet e instalar em seu computador²¹⁷. Assim como no caso do *Shell*, tive que baixar uma versão compatível com o Windows para que rodasse em meu computador. Infelizmente, essa versão não “conversa” com o Cygwin, ou seja, muitas funcionalidades do *Unix Shell*, que num computador Unix / Linux seriam ferramentas complementares do *Python*, não podem ser integradas às rotinas escritas neste programa no ambiente Windows. Para usar os comandos *Shell* nos dados obtidos numa rotina *Python*, é necessário abrir a janela Cygwin e escrever uma rotina em separado.

Não cabe aqui explicar todas as funcionalidades da linguagem *Python*. Na medida do possível (e quando julgar conveniente), reproduzirei as rotinas utilizadas nas etapas de compilação das listas de UTEs em português. Continuarei a usar o *corpus* de português para ilustrar os procedimentos; no caso da compilação, de fato, de um dicionário voltado para tradutores, todos os procedimentos de coleta de UTEs deverão ser repetidos no *corpus* de inglês.

4.2.5.1 Melhorando a etiquetagem do corpus

A qualidade da extração automática de padrões morfossintáticos depende, em muito, da acuidade da etiquetagem do corpus. Como foi mostrado no item 4.2.2.2, o etiquetador Tree-Tagger apresentou vários problemas ao atribuir a categoria morfossintática a itens lexicais do *corpus* de Culinária, em especial aqueles que têm um comportamento atípico em textos da área. Para solucionar esses problemas, o ideal seria corrigir as etiquetas erradas, manualmente, numa parte do corpus, incorporar essa informação ao programa e, então, etiquetar o *corpus* total novamente. Infelizmente, não se tem acesso ao repertório de regras de etiquetagem usados pelo programa Tree-Tagger (o arquivo em que estão registradas todas as palavras de seu “dicionário”, com as respectivas etiquetas a serem atribuídas, e como o programa deve proceder no caso de a palavra a ser etiquetada não constar desse dicionário) – teria que construir um repertório como esse do zero, caso quisesse treinar o etiquetador

²¹⁷ uma boa fonte para download do Python, especialmente para quem não está familiarizado com o programa, é: <http://nltk.sourceforge.net/index.php/Download>. Parte do site *Natural Language Tool Kit*, essa versão vem acompanhada de um tutorial e de um programa (IDLE – Integrated Development Environment) que auxilia o usuário a escrever e corrigir as rotinas a serem rodadas.

para trabalhar exclusivamente com textos culinários. Diante da dificuldade de execução e morosidade de tal tarefa, que também fugiria do escopo de desta pesquisa, optei por utilizar um etiquetador previamente treinado mas que permitisse melhorar a acuidade da etiquetagem por meio de correções pontuais.²¹⁸

Com base num *corpus* de língua geral de 184.985 palavras em português, previamente etiquetado e revisado manualmente por lingüistas (*corpus* Bosque²¹⁹, parte do projeto Floresta Sintá(c)tica²²⁰ da Linguateca), Baldrige criou um etiquetador que utiliza uma versão simplificada das etiquetas de categoria gramatical desse projeto²²¹:

ADJ	Adjetivo
ADV	Advérbio
ART	Artigo
CONJ	Conjunção
IN	Interjeição
NUM	Número Cardinal / Ordinal
N	Nome Comum / Próprio
PR	Pronome
PROP	Nome próprio
PRP	Preposição
V	Verbo
PUNC	Sinais de pontuação

O programa gera um arquivo que apresenta duas colunas separadas por marca de tabulação: uma como texto proveniente do *corpus* e uma com a categoria gramatical de cada palavra naquele contexto, conforme pode ser visto no fragmento a seguir:

```
uma      art
colher    v
de        prp
sopa      n
de        prp
maionese  n

cozinhe   v
a         art
batata    n
e         conj
```

²¹⁸ Para tanto, contei com o auxílio do Prof. Dr. Jason Baldrige, do Laboratório de Lingüística Computacional do Departamento de Lingüística da Universidade do Texas, em Austin, onde realizei o estágio de doutorado sanduíche.

²¹⁹ a versão utilizada foi a 7.3, disponível no endereço: http://nextens.uvt.nl/~conll/free_data.html

²²⁰ disponível em: <http://www.linguateca.pt/Floresta/>

²²¹ o conjunto de etiquetas completo pode ser visto no item 2.3 da página do projeto: <http://visl.sdu.dk/visl/pt/symbolset-floresta.html>

```

tempere      v
a            prp
gosto       n
.            punc

```

Utilizando um modelo estatístico de aprendizado denominado *Maximum Entropy Markov Model*²²², o etiquetador foi treinado, com base no *corpus* Bosque. Isto é, a partir das etiquetas revisadas constantes desse corpus, o modelo realiza uma série de cálculos estatísticos e cria regras a partir desses dados para que o etiquetador seja capaz de atribuir uma etiqueta a uma palavra de um *corpus* diferente, ainda que esta não conste de seu dicionário (ou seja, não faz parte da lista de palavras do *corpus* Bosque e, portanto, o programa não saberia, a princípio, que etiqueta atribuir-lhe).

Por exemplo: se a maioria das palavras do *corpus* Bosque que começam com letra maiúscula mas não estão no início de uma sentença são etiquetados como nomes próprios (prop), é possível que o modelo crie uma regra a partir dessa informação e, ao encontrar uma palavra com essas características e que não conste do *corpus* Bosque, atribua-lhe a etiqueta “prop”. Ou, se uma palavra tem mais de uma etiqueta nesse *corpus* (por exemplo, quadrado – substantivo / quadrado – adjetivo), serão observadas as frequências de cada uma, as etiquetas e/ou palavras que a precedem, etc. e, com base nessas informações, será calculada a probabilidade de uma e outra ocorrerem num texto diferente do usado no treinamento.

O grande diferencial de trabalhar com uma aplicação como essa é que a revisão manual de parte das etiquetas erradas surte grandes efeitos no processo de etiquetagem como um todo, pois o modelo estatístico, com base no trecho do *corpus* fornecido que contém as etiquetas corrigidas manualmente (chamado em computação de “golden collection”), refaz os cálculos de probabilidade, modificando suas regras – ele “reaprende” a etiquetar a cada novo trecho revisado que é incorporado ao seu dicionário.

Infelizmente, não é simples implementar ferramentas como essa, pois a maioria das aplicações não possui interface amigável para quem não tem conhecimentos aprofundados de computação²²³. Idealmente, esta etapa do trabalho de levantamento

²²² a aplicação desse modelo a etiquetadores foi descrita pela primeira vez em Ratnaparkhi, 1996 (disponível em: <http://acl.ldc.upenn.edu/W/W96/W96-0213.pdf>)

²²³ para implementar esta e outras aplicações computacionais de processamento da linguagem natural desenvolvidas na Universidade do Texas em Austin, visite a página do projeto TextNLP (Text Natural Language Processing): <http://comp.ling.utexas.edu/software/textnlp>.

de UTEs deve ser feita em parceria com algum pesquisador da área de processamento da linguagem natural ou computação. Caso isso não seja possível, pode-se usar um etiquetador como o Tree-Tagger e corrigir o *corpus* etiquetado diretamente, usando substituições em massa – o que, como veremos mais adiante, acaba sendo uma etapa inevitável.

Para etiquetar o *corpus* de Culinária (português) usando o etiquetador desenvolvido por Baldridge, enviei o arquivo contendo todos os textos em formato somente texto (.txt) e itemizado (uma palavra por linha) que havia usado com o etiquetador Tree-Tagger. Em posse dos resultados, observei, como esperava, que o etiquetador apresentava vários problemas na análise dos itens lexicais presentes nas receitas culinárias, entre eles:

1. Considerar a primeira palavra de uma sentença, em letras maiúsculas (geralmente um verbo, no modo de fazer das receitas) como nome próprio (etiqueta “prop”):

Retire	prop	Aque a	prop
do	v	numa	adj
forno	n	frigideira	n
Mantenha	prop	Retire	prop
o	art	do	v
forno	n	fogo	n

2. Considerar adjetivos deverbais como verbo:

miolo	n	
de	prp	
cebola	n	
reservado		v
bem	adv	
picado		v

3. Transformar todos os caracteres acentuados, cedilhas e frações em um único símbolo (que aparece como uma interrogação no corpus, mas, quando lido num arquivo somente texto do tipo ANSI, transforma-se numa seqüência longa e confusa – “\xef\xbf\xbd”) que, além de impedir sua substituição em massa, faz com que o programa etiquete erroneamente as palavras que contém tal caractere:

Ligue prp
 o art
 forno n
 punc
 temperatura n
 de prp

4. Etiquetar como “substantivo” todas as ocorrências de verbo + pronome oblíquo:

coloque-o n
 em prp
 uma art

misture-o n
 com prp

O passo seguinte foi, portanto, melhorar a qualidade da etiquetagem – processo trabalhoso e demorado, mas crucial para o bom desempenho da extração semi-automatizada baseada em padrões morfossintáticos. Numa primeira etapa, revisei manualmente três trechos aleatórios do corpus, um do começo, um do meio e um do fim do arquivo (cerca de 15 receitas inteiras – aproximadamente 1.400 palavras, no total). Essa informação foi incorporada ao dicionário e o etiquetador foi treinado novamente pelo Prof. Baldrige. A nova versão etiquetada do *corpus* me foi enviada.

Muitos problemas foram resolvidos, ainda que parcialmente (como a etiquetagem errônea de todos os verbos em início da sentença como nome próprio – “prop”, devido à letra maiúscula). Mas, a irregularidade na formatação dos textos do corpus, segundo Baldrige, foi o fator responsável pela grande maioria dos erros do etiquetador, que tem seu desempenho ótimo apenas quando trabalha com arquivos em que há uma linha em branco, e apenas uma, separando as sentenças do corpus. Assim, numa segunda fase da revisão, me propus reparar essa falha de forma automática – uma etapa demorada e laboriosa.

Depois de muitas tentativas frustradas de corrigir os espaços em branco, marcas de parágrafo e de tabulação excedentes do *corpus* usando o *Shell Unix*, o *Python*, o Text Converter do WST e até mesmo o Substituir (Find & Replace) do *Word*, não consegui corrigir todos os arquivos. Por esse motivo, acabei por eliminar uma parte do *corpus* de português deste estudo (cerca de 275.708 palavras: os subcorpora CDrec, JPan e espec_PO – vide item 3.5.3.2), pois não consegui controlar a quantidade de linhas em branco entre as sentenças nas receitas desses arquivos automaticamente.

O novo *corpus* reformatado foi enviado, pela última vez, para ser etiquetado. O resultado foi estudado com buscas aleatórias usando as ferramentas do WST. A partir desse ponto, os erros passaram a ser corrigidos diretamente no *corpus*, usando os procedimentos detalhados a seguir.

Escrevi, primeiramente, uma rotina em *Python* (*words_with_more_than_1_tag.sh*) para identificar todas as palavras do novo *corpus* que tivessem recebido mais de uma etiqueta – erros potenciais. As palavras etiquetadas como substantivo e como nome próprio (por exemplo: *abacaxi n* e *Abacaxi prop*) foram gravadas num arquivo à parte.

```
# Elisa D. Teixeira, 05/Dec/2007
#
# Given a filename (which must be in the same directory of the routine),
# the program inspects "word    part of speech" pairs, stores occurrences
# and at the end print out all words that occur with more
# than one part of speech

f = open("C:/Python25/PhD_scripts/Jason_tagged/Jason_tagged.txt", "r")
output1 = open("more_than_1_POS_Jason_tagged.txt", "w")
output2 = open("more_than_1_POS_nom_prop.txt", "w")
word_2_pos = {}
final_word_2_POS = {}
import string

for line in f:
    for word in line.split():
        pieces = word.split("\t")
        if len(pieces) < 2:
            # We're expecting _every_ word in the file to have the form
            # "word    POS". If we don't find a "\t", then there's something
            # wrong
            print "problem in: ", pieces
            pass
        # recover "word    POS" pairs and remember the POS seen for each word
        else:
            bareword = pieces[0]
            pos = pieces[1]
            if word_2_pos.has_key(bareword):
                if pos not in word_2_pos[bareword]:
                    word_2_pos[bareword].append(pos)
            else:
                word_2_pos[bareword] = [ pos ]

# print words with more than one POS
for word in word_2_pos.keys():
    if len(word_2_pos[word]) > 1:
        if len(word_2_pos[word]) == 2:
            if "n" in word_2_pos[word] and "prop" in word_2_pos[word]:
                print >> output 2, word
            else:
                print >> output1, word, ":", word_2_pos[word]
        else:
            print >> output1, word, ":", word_2_pos[word]
output.close()
```

Para que o *Python* execute a rotina, basta clicar, no menu da tela do IDLE (IDLE – Integrated Development Environment, o editor de rotinas para Python que acompanha o pacote do programa no NLTK – Natural Language Tool Kit²²⁴): *Run > Run Module* (ou apertar a tecla de atalho F5). Como é possível ver na Fig. 68, esse editor faz as indentações automaticamente na rotina, além de trabalhar com um sistema de cores – por exemplo, palavras-chave do programa em laranja, linhas ignoradas (precedidas de “#”) em vermelho – que facilita a visualização dos comandos e a redação do programa.

Figura 68: tela do editor IDLE e *prompt* do programa *Python*

Essas palavras e respectivas etiquetas foram gravadas num arquivo somente texto que, em seguida, foi ordenado alfabeticamente usando o *Unix Shell*. O arquivo resultante tem a seguinte aparência²²⁵:

```
: ['punc', 'art', 'n', 'prp', 'num', 'adv', 'adj', 'v', 'prop',
'pron']
150 C : ['num', 'n', 'punc']
```

²²⁴ <http://nltk.sourceforge.net/index.php/Download>

²²⁵ como pode ser observado, a presença desse símbolo ocasiona vários erros de etiquetagem

```

160 C : ['n', 'num']
170 C : ['num', 'n']
175 C : ['num', 'n']
180 C : ['punc', 'n', 'num']
190 C : ['num', 'n']
200 C : ['num', 'n']
220 C : ['num', 'n']
3 : ['n', 'num']
a : ['v', 'art', 'prp']
a a : ['n', 'v']
A a : ['v', 'prop']
a afr o : ['n', 'adj']
ab bora : ['n', 'adj', 'art', 'v']
ab bora-moranga : ['n', 'v']

```

As etiquetas foram revisadas manualmente (565 palavras, no total). Em seguida, corriji manualmente as 2.771 etiquetas do arquivo que continha apenas as palavras que receberam as etiquetas de substantivo e nome próprio. Por fim, substitui, diretamente no corpus, todas as ocorrências erradas pela etiqueta certa. A rotina usada para executar essa tarefa (*wrong_tag_reassign.sh*) foi:

```

# Elisa D. Teixeira, 19/Oct/2007

# this routine substitutes wrong tags based on a file with the correct tags
# for those words with more than 1 tag

filename1 =
open("C:/Python25/PhD_scripts/Jason_tagged/Jason_corrected_phase1.txt", "r")
corpus_wrong_tags = filename1.read()
filename2 = open("Jason_corrected_tags.txt","r")
correcttags = filename2.read()
corpus_correct =
open("C:/Python25/PhD_scripts/Jason_tagged/Jason_corrected_phase2.txt", "w")
import string
import re

tagdict = {}
for item in correcttags.split():
    word,tag = item.split("\t")
    tagdict[word] = tag

for wtpair in corpus_wrong_tags.strip().split(" "):
    slash_index = wtpair.rfind("\t")
    w,t = wtpair[:slash_index],wtpair[slash_index+1:]
    if w in tagdict.keys():
        print >> corpus_correct, w+"\t"+tagdict[w],
    else:
        print >> corpus_correct, wtpair,

```


No passo final da correção da etiquetagem, foram feitas algumas substituições em massa, por exemplo, foi acrescentado um espaço em branco antes das medidas (*g, gr, cm, l, kg, °C, h, m, min, mm, ml* e variações), separando-as dos números, e um espaço antes e depois das barras (mas ignorando as frações), para separar itens como “médio/baixo”, “manteiga/margarina”, etc. Depois foram feitas substituições baseadas em regras, criadas a partir da observação das linhas de concordância de palavras problemáticas, para as que podiam ter mais de uma etiqueta. Por exemplo:

- “grosso”:
 - se for precedido de verbo = adv
 - se não for = adj
- “mesmo”:
 - se for precedido de artigo = n
 - se for precedido de preposição = adj
 - todos os outros casos = deixar como está
- “clara”:
 - se precedido de numeral, artigo, preposição, pronome ou sinal de pontuação = n
 - se precedida de substantivo ou advérbio = adj
 - se precedida de conjunção – se houver gema no contexto imediato = n / se não houver = adj

Uma vez minimizados os principais problemas de etiquetagem do corpus, passei à extração dos padrões morfossintáticos identificados no estudo das UTEs provenientes da memória de tradução. Para facilitar as buscas e a visualização dos resultados, substituí as marcas de tabulação do arquivo por uma barra ("/"), eliminei as marcas de parágrafos usando uma rotina em *Unix Shell* – `tr -s '\t' '/' < nome_do_arq_com_tabs.txt | tr -d '\r' | tr -s # '\n' ' ' > nome_do_arquivo_sem_tabs.txt` – e substituí os caracteres especiais (“\xef\xbf\xbd”) por um sinal de interrogação usando o aplicativo *Substituir (Find & Replace)* do próprio Bloco de Notas. A aparência final do *corpus* etiquetado ficou assim:

```
COUSCOUS/prop DE/prp COLIFLOR/prop INGREDIENTES/prop (/punc 4/num
Por??es/n )/punc 1/num couve-flor/n pequena/adj 50/num g/n de/prp
p?o/n de/prp especiarias/n com/prp mel/n 30/num ml/n de/prp ?leo/n
de/prp oliva/n 300/num ml/n de/prp Campari/prop 4/num g/n de/prp agar-
agar/v em/prp p?/n 100/num ml/n de/prp vinagre/n de/prp Cabernet/prop
Sauvignon/prop 5/num g/n de/prp a??car/n 2/num g/n de/prp glicose/n
Sal/n a/art gosto/n MONTAGEM/prop 4/num folhas/n de/prp hortel?/n
4/num ramos/n de/prp (...)
```

4.2.5.2 Extração das frequências por categoria inicial da UTE

Para recuperar as UTEs no *corpus* usando os padrões morfossintáticos previamente identificados, foi necessário, primeiramente, transformar esses padrões em expressões regulares. Uma expressão regular, chamada também de “regex” (acrônimo de *regular expression*) em computação, usa uma série de caracteres especiais (ou “metacaracteres”) para representar elementos textuais de forma a permitir buscas por padrões²²⁶. Por exemplo, o caractere “?”, quando usado numa expressão regular, significa: 0 ou mais ocorrências de um caractere do mesmo tipo do que precede a interrogação. Assim, uma busca no *corpus* por “colo?” recupera as ocorrências de “color”, “colour”.

No caso dos padrões morfossintáticos que se pretende extrair do corpus, uma estrutura simples de busca para – por exemplo, substantivo + preposição + substantivo (“n + prp + n” no *corpus* de estudo etiquetado) – seria:

`\s+(\S+)/n\s+(\S+)/prp\s+(\S+)/n\s+`, em que:

- `\s` = espaço em branco, marca de tabulação ou quebra de linha
- `+` = uma ou mais ocorrências
- `\S` = tudo que não for espaço em branco (que também pode ser escrito: `^\s`)
- `( )` = delimita a porção do resultado que deve ser exibido (neste caso, retorna apenas as palavras, e não as etiquetas)
- `/n`, `/prp`, `/n` = etiquetas do *corpus* para substantivo, preposição e substantivo, respectivamente

Assim, para cada tabela do item 4.2.2.3 foi gerado um arquivo somente texto (*NOM_regex_patterns.txt*, *VERB_regex_patterns.txt*, *ART_regex_patterns.txt*, *ADV_regex_patterns.txt*, *PRP_regex_patterns.txt*, *ADJ_regex_patterns.txt*) contendo as

²²⁶ para maiores informações sobre como construir expressões regulares, vide: <http://www.regular-expressions.info/quickstart.html>

expressões regulares correspondentes, uma após a outra, separadas por um espaço em branco (não pode haver um espaço em branco ao final da última expressão). Alguns padrões morfossintáticos tiveram que ser alterados para se adequarem às categorias usadas pelo etiquetador de Baldrige, pois eram diferentes das usadas pelo Tree-Tagger.

Em seguida, escrevi uma rotina em Python para fazer as buscas no *corpus* (*search_regex_patterns.sh*):

```
# Elisa Duarte Teixeira, 07/Nov/2007

# this routine takes as input a tagged file (word/tag format) and performs
# searches for a list of tag patterns, previously transformed into Regular
# expressions and saved, separated by a single space, in a txt file

# directory, folder and name of the tagged file
x = open("C:/Python25/PhD_scripts//corpus/corpus_tese_pbtt.txt", "r")
tagtext = x.read()
# file containing regular expressions separated by single space
y = open("NOM_regex_patterns.txt", "r")
re_patterns = y.read()
# file where the results are to be stored
output = open("NOM_found_pattern_PBTt.txt", "w")
import re
import string
import os
# previous line does not work within windows/cygwin, but it does in Unix /
# Linux, allowing Shell commands to be used within Python - just erase "#"

for expr in re_patterns.split(" "):
    UTEmatch = re.findall(expr, tagtext)
    print >> output, "# searched regular expression: ", expr
    print >> output, "\n"
    for item in UTEmatch:
        print >> output, item
    print >> output, "\n"
output.close()

# deletes everything that is not letters and whitespaces to obtain a list
# again and then eliminates repeated patterns, counting the occurrences.

# if using cygwin:
# tr -Cs "A-Za-z0-9\n\\#\\\" \" \" < NOM_found_pattern_PBTt.txt | sort | uniq -c
# | sort -nr > N_patternd_freq_PBTt.txt
```

O programa gera um arquivo do tipo texto com os resultados que tem a seguinte aparência:

```

# searched regular expression: \s+(\S+)/art\s+(\S+)/n\s+(\S+)/adj\s+

('as', 'folhas', 'externas')
('um', 'rel?gio', 'imagin?rio')
('uma', 'massa', 'grossa')
('o', '?leo', 'extravirgem')
('o', '?leo', 'extravirgem')
('o', 'molho', 'morno')
('um', 'molho', 'grosso')
('o', 'vinho', 'branco')
('o', '?leo', 'extravirgem')
('um', 'p?', 'fino')
('o', 'ch?', 'c?trico')
('os', 'cogumelos', 'secos')
('uma', 'peneira', 'fina')
('a', 'flor', 'inteira')
('os', 'alhos', 'amassados')
('a', 'cebolinha', 'verde')
('uma', 'panela', 'limpa')
('o', 'alho', 'amassado')
('o', 'sal', 'grosso')
('as', 'ovas', 'pretas')
('a', 'manteiga', 'restante')
('a', 'rodela', 'central')
('uma', 'mistura', 'cremosa')
('as', 'f?rmas', 'preparadas')
('uma', 'vasilha', 'grande')
('uma', 'assadeira', 'untada')

```

Figura 69: aparência do arquivo gerado pelo programa

Fiz um exame rápido dos dados obtidos para ver, por exemplo, se havia padrões com um número aparentemente anormal de ocorrências. Uma vez corrigidos os erros na rotina e nas expressões regulares – fase extremamente trabalhosa e demorada – preparei os arquivos para que os resultados pudessem ser analisados mais facilmente. Como o *Python* e o *Cygwin* não interagem no sistema operacional Windows, usou-se o *Cygwin*, em separado, para agrupar e contar as ocorrências repetidas com a rotina:

```
tr -Cs "A-Za-z0-9\n\#\\" " " < arq_gerado_pelo_Python.txt | sort
| uniq -c | sort -nr > arq_com_UTEs_e_frequencia.txt
```

O resultado foi colado num arquivo Word, os padrões com ocorrência abaixo de 3 foram eliminados e, após uma série de substituições em massa usando a ferramenta *Substituir (Find & Replace)* do Word (basicamente eliminando espaços excedentes e colocando um sinal de tabulação para separar cada expressão de sua frequência), o resultado foi colado numa planilha do Excel que contém duas colunas: uma com as UTEs e uma com sua respectiva frequência no corpus:

	A	B	C	D	E	F	G	H	I	J	K
1	colheres sopa	6930									
2	x?cara ch?	3841									
3	Modo de Preparo	3835									
4	colher sopa	3370									
5	Tempo de preparo	2439									
6	colher ch?	2322									
7	azeite de oliva	2077									
8	farinha de trigo	1930									
9	sal e pimenta do reino	1899									
10	?gua na Boca Categoria	1527									
11	?gua na Boca	1527									
12	Boca Categoria	1527									
13	Revista ?gua	1510									
14	suco de lim?o	1379									
15	creme de leite	1312									
16	gosto Modo	1188									
17	x?caras ch?	1105									
18	dentes de alho	1101									
19	min Rendimento	1075									
20	Categoria Sobremesas	1015									
21	colheres de sopa	989									
22	sal e pimenta	953									
23	fogo baixo	920									
24	vinho branco	845									
25	por??es Calorias	842									
26	?gua fervente	656									

Figura 70: Tabela do Excel para abrigar padrões recuperados

À parte, escrevi uma rotina em *Python* para saber a ocorrência total no *corpus* de cada padrão individual, já que a rotina escrita previamente contava apenas os padrões com ocorrência acima de 3:

```
# Elisa Duarte Teixeira, 07/Nov/2007

# this routine counts the occurrences of predefined patterns in a
# tagged corpus

# name of the tagged file
x = open("C:/Python25/PhD_scripts//corpus/corpus_tese_pbtt.txt", "r")
tagtext = x.read()
y = open("NOM_regex_patterns.txt", "r")
# file containing regular expressions separated by single space
re_patterns = y.read()
print re_patterns.split(" ")
# file with results
output = open("freq_N_patterns.txt", "w")
import re
import string

for expr in re_patterns.split(" "):
    UTEmatch = re.findall(expr, tagtext)
    print >> output, expr
    counter = 0
    for item in UTEmatch:
        counter = counter + 1
    print >> output, counter
    print >> output, "\n"
output.close()
```

Prossigui com a pesquisa selecionando amostras aleatórias para análise da precisão (procedimento que detalho a seguir, no item 3.2.5.4). No entanto, no momento de quantificar os dados, me dei conta de que teria sido muito mais vantajoso (e muito mais fácil para a análise) se tivesse escrito um programa que já contasse as ocorrências no *corpus* e as agrupasse por expressão de busca, eliminando, portanto, a fase de agrupamento e ordenação dos resultados com a rotina Cygwin – que acaba por dissociar os resultados das expressões de busca que os originaram ao colocar todas as linhas do arquivo em ordem alfabética. Ou seja, não seria possível saber, de pronto, qual expressão de busca originou aquele resultado – na Fig. 70, por exemplo, é possível observar que “Modo de Preparo” ocorreu 3.835 vezes no corpus, mas não é possível saber qual expressão de busca recuperou essas ocorrências. Em vista disso, uma nova rotina foi escrita, que já calcula todos os dados numéricos: total de ocorrências do padrão no corpus, total de formas distintas (types) e total de ocorrências geradas por esses types com ocorrência maior ou igual a 3:

```
# Elisa Duarte Teixeira, 30/Jul/2008

# this routine takes as input a tagged file (word/tag format) and performs
# searches for a list of tag patterns, previously transformed into Regular
# expressions and saved, separated by a single space, in a txt file

# directory, folder and name of the tagged file
x = open("C:/Python25/PhD_scripts//corpus/corpus_tese_pbtt.txt", "r")
tagtext = x.read()
# file containing regular expressions separated by single space
y = open("NOM_regex_patterns.txt", "r")
re_patterns = y.read()
# file where the results are to be stored
output = open("NOM_found_pattern_PBTt.txt", "w")
import re
import string

# for each regex, find matching cases in the corpus, store and count them
# and then print only those with frequency equal or greater than 3. Also
# counts how many "types" there are in the list with matching expressions
# occurring more than twice in the corpus

for expr in re_patterns.split(" "):
    regex_count = {}
    counter1 = 0
    counter2 = 0
    counter3 = 0
    counter4 = 0
    UTEmatch = re.findall(expr, tagtext)
    print >> output, "# searched regular expression: ", expr
    print >> output, "\n"
    for item in UTEmatch:
        regex_count[item]=regex_count.get(item, 0) + 1
        counter1 += 1
    for item in regex_count.keys():
```

```

counter4 += 1
print >> output, item, "\t", regex_count[item]
if regex_count[item] > 2:
    counter2 += 1
    counter3 = counter3 + regex_count[item]
else:
    pass
print >> output, "\n"
print >> output, "total de oc. do padrão no corpus: ", "\t", counter1
print >> output, "total de formas distintas (types): ", "\t",
counter4
print >> output, "total de formas com oc. > 2: ", "\t", counter2
print >> output, "total de oc. dos types com oc > 2: ", counter3
print >> output, "\n"
output.close()

```

Os resultados foram colados numa nova tabela do Excel, que contém uma coluna para abrigar o padrão gramatical que originou a expressão recuperada²²⁷:

	A	B	C	D	E	F	G	H
		freq	padrão pesquisado	UTE?				
1	forma recuperada							
2	colheres sopa	6922	n + punc + n + punc					
3	Modo de Preparo	3835	n + prp + n					
4	x?cara ch?	3833	n + punc + n + punc					
5	colher sopa	3363	n + punc + n + punc					
6	Tempo de preparo	2439	n + prp + n					
7	colher ch?	2321	n + punc + n + punc					
8	azeite de oliva	2077	n + prp + n					
9	farinha de trigo	1930	n + prp + n					
10	?gua na Boca	1527	n + prp + n					
11	?gua na Boca Categoria	1527	n + prp + n + n					
12	Boca Categoria	1527	n + n					
13	sal e pimenta-do-reino	1517	n + conj + n					
14	Revista ?gua	1510	n + n					
15	suco de lim?o	1379	n + prp + n					
16	creme de leite	1312	n + prp + n					
17	gosto Modo	1188	n + n					
18	x?caras ch?	1104	n + punc + n + punc					
19	dentes de alho	1101	n + prp + n					
20	Categoria : Sobremesas :	1015	n + punc + n + punc					
21	colheres de sopa	989	n + prp + n					
22	sal e pimenta	952	n + conj + n					
23	fogo baixo	920	n + adj					
24	vinho branco	845	n + adj					
25	por??es Calorias	842	n + n					

Figura 70: aparência da nova tabela Excel com os resultados

4.2.5.3 Frequência dos padrões no *corpus* de Culinária total

²²⁷ nos padrões **n + punc + n + punc**, “punc” refere-se aos parênteses, p. ex.: “colheres (sopa)”

A extração de candidatas a UTE no *corpus* de português de 1,5 milhões de palavras, a partir dos padrões observados na lista de UTEs proveniente da memória de tradução, gerou um grande número de ocorrências. A tabela 16 mostra os dados quantitativos gerais dos padrões pesquisados de acordo com a categoria inicial da expressão de busca explicitando: o número de padrões pesquisados, a frequência geral no corpus, o número de formas distintas (*types*) recuperadas e quantas delas tiveram ocorrência maior ou igual a 3, com respectivas percentagens.

Por exemplo: o padrão **v + art + n + prp + art + n + adj + prp + n**, originário da UTE “*abra a massa sobre uma superfície polvilhada com farinha*” (da memória de tradução), recuperou 42 formas no corpus, isto é, essa sequência exata de categorias gramaticais ocorreu 42 vezes no corpus. Se observarmos a lista de formas resultante, veremos que há ocorrências repetidas (de palavras, e não mais de categorias gramaticais): “*abra a massa sobre uma superfície polvilhada com farinha*” ocorreu 2 vezes no *corpus* (= 1 *type*), e a expressão “*Coloque a massa sobre uma superfície polvilhada com farinha*” ocorreu 4 vezes (= 1 *type*). Todas as demais ocorreram apenas uma vez no corpus. Assim, esse padrão teve: ocorrência total no *corpus* = **42**; formas distintas (*types*) = **38**; no. de *types* com ocorrências maior ou igual a 3 = **1**. Esse procedimento foi feito para cada uma das 130 expressões pesquisadas, e os resultados foram agrupados, na tabela a seguir, pela categoria inicial do padrão de busca (os dados detalhados de cada expressão pesquisada, discutidos mais adiante, podem ser vistos no Anexo V).

padrão pesquisado (vide tabelas 8 a 14)	no. de padrões pesq.	oc. total no corpus	<i>types</i> (% do total de formas recuperadas)	<i>types</i> freq. ≥ 3 (% total / % das formas distintas)
iniciados por substantivo (n)	35	262.527	68.731 (26%)	11.153 (4% / 16%)
iniciados por preposição (prp)	27	234.343	23.184 (10%)	6.395 (3% / 28%)
iniciados por verbo (v)	40	145.855	55.545 (38%)	8.009 (5% / 14%)
iniciados por artigo (art)	4	32.602	10.551 (32%)	1.914 (6% / 18%)
iniciados por adjetivo (adj)	11	15.214	7.056 (46 %)	1.025 (7% / 15%)
iniciados por advérbio (adv)	13	13.588	3.766 (28%)	838 (6% / 22%)
Totais	130	704.129	168.833 (24%)	29.334 (4% / 17%)

Tabela 16: dados quantitativos gerais sobre a extração por padrões

Os 130 padrões pesquisados foram encontrados 704.129 vezes no *corpus* de português. No total, foram recuperadas 168.833 formas distintas (*types*) – cerca de 25% do total de formas recuperadas – e, dentre essas, 29.334 tiveram ocorrência maior ou igual a 3 no *corpus* (cerca de 17 % do total de *types*, 4% de todas as formas recuperadas). Esses 4% (cerca de 30.000 formas) seriam, a princípio, o objeto de análise para avaliar a precisão da extração por padrões morfossintáticos.

Os padrões iniciados por substantivo e preposição são responsáveis por dois terços dessas ocorrências (um terço cada), seguidos pelos iniciados por verbo, artigo, adjetivo e advérbio. O gráfico da figura 71a mostra a proporção geral de unidades recuperadas no *corpus* total por categoria inicial do padrão de busca.

O gráfico da figura 71b mostra, dentre o total de formas recuperadas (eixo y – somatória de todas as cores da coluna), quantas são formas distintas (*types* – somatória das cores vermelho e verde) e, dentre estas, quantas tiveram ocorrência igual ou maior que 3 no *corpus* (cor verde), por categoria inicial pesquisada.

Como é possível observar, ainda que o maior número de padrões pesquisados seja iniciado por verbos (= 40), foram os iniciados por substantivos (= 35) que renderam mais formas distintas, o que poderia ser atribuído à maior variedade lexical dos substantivos da área (ingredientes, equipamentos, preparações) se comparados à variedade de verbos (processos). Chama a atenção também a proporção de unidades com ocorrência igual ou maior que 3 dentre os padrões iniciados por preposição e advérbio (mais de 20%) – um indício de que tais categorias, ditas “fechadas”, apresentam não apenas variedade combinatória (28% dos padrões iniciados por advérbio recuperados são formas distintas), mas uma frequência expressiva em receitas. A categoria artigos, com apenas 4 expressões de busca, gerou uma lista de 32.602 padrões (6% dos quais têm frequência maior ou igual a 3 no *corpus*).

Figura 71a: total geral de padrões recuperados por categoria

Figura 71b: proporção de formas distintas (types) ocorrendo 3 ou mais vezes

Se o número de ocorrências recuperadas apenas pelos *types* com ocorrência maior ou igual a três no *corpus* (vide Tab. 17) for analisado, observa-se que o grau de repetição é muito alto: os 11.153 *types* com ocorrência maior ou igual a 3 iniciados por substantivo (que representam 16% de todos os *types* recuperados para essa categoria inicial) são responsáveis por 195.678 ocorrências (*tokens*) – 75% de todas as

ocorrências dessa categoria no corpus. Percentuais semelhantes foram obtidos pelas categorias iniciadas por advérbios (74%) e verbos (63%). Os iniciados por preposição chamam a atenção por seu alto percentual de *types* com oc. ≥ 3 (92%), ao passo que as iniciadas por adjetivo e artigo destacam-se das demais categorias por apresentarem um percentual mais baixo (54% e 47%, respectivamente).

Seriam UTEs todas essas 6.395 formas iniciadas por preposição recuperadas pelos 40 padrões originados da memória de tradução? Quantas dessas 29.334 formas recuperadas, que totalizaram 534.776 ocorrências no *corpus* de culinária português, afinal, poderiam ser consideradas Unidades de Tradução Especializada?

padrão pesquisado (vide Tabela X)	oc. total no corpus	<i>types</i> c/ freq. ≥ 3 (%) do total <i>types</i>)	ocs. geradas p/ <i>types</i> c/ freq. ≥ 3 (% oc. tt)
iniciados por substantivo (n)	262.527	11.153 (16%)	195.578 (75%)
iniciados por preposição (prp)	234.343	6.395 (28%)	214.472 (92%)
iniciados por verbo (v)	145.855	8.009 (14%)	91.231 (63%)
iniciados por artigo (art)	32.602	1.914 (18%)	15.214 (47%)
iniciados por adjetivo (adj)	15.214	1.025 (15%)	8.257 (54%)
iniciados por advérbio (adv)	13.588	838 (22%)	10.024 (74%)
Totais	704.129	29.334 (17%)	534.776 (76%)

Tabela 17: proporção entre formas distintas (*types*) e frequência (*tokens*)

Para responder a essa pergunta, o passo seguinte da pesquisa consistiu em avaliar a precisão da metodologia de extração empregada. Diante da impossibilidade de se analisar todos os resultados obtidos, foi feita uma análise de amostras aleatórias – procedimento detalhado no subitem a seguir.

4.2.5.4 Análise da precisão em amostras aleatórias

Toda vez que o pesquisador se depara com uma quantidade de dados grande demais para ser analisada manualmente – etapa imprescindível na coleta de dados usando quaisquer metodologias de extração (semi-)automáticas – o ideal é fazer uma análise de uma amostra que, preferencialmente, deve ser selecionada aleatoriamente. Para selecionar uma amostra aleatória de uma lista de itens como a com que estou trabalhando pode-se usar, por exemplo, a função “Random” do programa Excel. Para tanto, é preciso primeiramente criar uma coluna na tabela à esquerda dos dados para

abrigar o número randômico que o programa vai gerar para cada linha. Em seguida, clica-se na primeira célula desta coluna e digita-se, na janela dos dados, a função que gera o número randômico: “=RANDBETWEEN(1;no. de linhas da tabela)” (sem as aspas e substituindo a expressão grifada pelo número). Ao ser confirmado o valor para a célula, o programa inscreve um número aleatório entre 1 e o número total de linhas da tabela naquela célula (vide setas 1 e 2 na Fig. 72). Em seguida, ilumina-se todas as linhas válidas desta coluna, inclusive a que contém o número randômico gerado pelo programa, e clica-se sobre a tecla do menu “Editing” que preenche todas as células usando a mesma fórmula (seta 3 da Fig. 72; atalho: Ctrl+Down). O programa vai colocar um número aleatório em cada uma das células.

Por fim, reordenam-se as linhas pela coluna que contém os números randômicos por ordem crescente (menu Editing – seta 4 na Fig. 72) e seleciona-se a amostra (por exemplo, as 1.000 primeiras linhas). As linhas da amostra escolhida são coladas numa nova planilha antes de se começar a trabalhar e a coluna com os números randômicos é eliminada.

Figura 72: usando a função “Random” do Excel

Para selecionar as amostras randômicas para o presente estudo, esse método foi aplicado às diversas tabelas do Excel que havia criado para comportar as unidades

com frequência maior ou igual a 3 iniciadas por uma dada categoria (substantivo, por exemplo) e selecionei uma parte dessas ocorrências (pelo menos 20% do total de *types*). O Anexo V apresenta uma tabela com os dados completos da coleta por padrão pesquisado contendo: sua frequência geral no corpus, o número de formas recuperadas (*types*) com ocorrência maior ou igual a 3, o número de ocorrências que essas formas geraram (*tokens*) e alguns exemplos de UTEs que foram recuperados por cada padrão e de formas que não considere como UTE. O Anexo VI apresenta uma lista com as 100 UTEs mais recorrentes de cada categoria inicial identificadas nas amostras randômicas analisadas, em ordem alfabética.

Os resultados obtidos na análise das amostras aleatórias foram:

categoria inicial	types oc. ≥ 3	nº. de types analisados (% das oc. ≥ 3)	porcentagem de UTEs válidas
iniciados por substantivo (n)	11.154	2.500 (23%)	35%
iniciados por preposição (prp) - com <i>stoplist</i> ²²⁸	3.502	1.000 (28%)	25%
iniciados por verbo (v)	5.101	1.000 (20%)	50%
iniciados por artigo (art)	1.915	500 (26%)	25%
iniciados por adjetivo (adj)	1.026	1.026 (100%)	35%
iniciados por advérbio (adv)	827	300 (37%)	48%
Totais	23.525	6.326 (27%)	36%

Tabela 18: quadro-resumo dos resultados da análise das amostras aleatórias

Nos itens que se seguem, são discutidos os resultados da análise das amostras e sugeridas melhorias para a metodologia de extração.

A. Padrões iniciados por substantivo

Na extração de padrões iniciados por substantivo, o que mais chama a atenção, à primeira vista, é a enorme ocorrência de medidas (*xícara(s)*, *colher(es)*, *colh.*, *litro(s)*, *l*, *ml*, *kg*, *g*, *cm*, *mm*, *copo*, etc.) e quantificadores (*pitada*, *pouco*, *cubo*, *pedaç(o/inho)*, *garrafa*, *dose*, *fatia*, *metade*, etc.), em geral seguidas pela preposição “de” (além de “pedaços” de medidas: *chá de*, *sopa de*, *café de*, etc.) na posição inicial dos padrões recuperados. Em sua grande maioria, essas unidades não constituem UTEs, embora possam ser usadas, quando ordenadas alfabeticamente, para estabelecer relações

²²⁸ vide discussão sobre os padrões iniciados por preposição (item B), logo após a tabela.

semânticas entre medidas e ingredientes específicos (por exemplo, quais ingredientes são usualmente medidos em *litros*, *pitadas*, *pedaços*, *vidros*, *latas*, etc.?; quais são cortados em *fatias*, *em rodela*s, *em cubos*?). Mas, do ponto de vista da extração automática, seria mais vantajoso criar uma *stoplist* para ignorar as ocorrências iniciadas por esses itens lexicais – que poderiam ser pesquisados posteriormente com ferramentas mais adequadas ao estudo das relações semânticas.

Com relação ao padrão “n + prp + n”, que é parte de outros padrões maiores, observei que as formas que contêm “preposição + artigo definido” = (*do*, *da*, *dos*, *das*, *num*, *numa*, *no*, *na*, *nos*, *nas*, *ao*, *aos*) ou em que a preposição “*com*”, de modo geral, não renderam UTEs. Seria vantajoso, portanto, reescrever os padrões (ou escrever uma rotina em Python) para que as ocorrências recuperadas contendo essas preposições na posição “n + PRP + n” fossem ignoradas. Observei também que os padrões que contêm “n + adj” teriam maior precisão na extração caso eu ignorasse alguns adjetivos de alta ocorrência no *corpus* e que, no geral, não formam UTEs (p. ex.: *grande*, *pequeno*, *médio*, *picado*, *fatiado*).

Outra característica geral observada nesses padrões é a grande presença de títulos de receitas (ou parte deles) na porção “n + prp” dos padrões recuperados, especialmente os que se referem a preparações mais comuns (*molho de*, *creme de*, *salada de*, *sanduíche de*, *lasanha de*, *omelete de*, *quiche de*, *pudim de*, etc.). Uma maneira eficiente de eliminar esse ruído nas listas de candidatas a UTEs seria etiquetar o título – bem como outras seções das receitas que interferiram negativamente nos resultados da extração, como as informações nutricionais, categoria, grau de dificuldade, etc. – para que possam ser excluídos das buscas (procedimento que havia sido feito inicialmente em parte do *corpus*, mas abandonado depois por não ter sido etiquetar automaticamente o *corpus* todo – vide item 3.5.3.1 para mais detalhes).

Outro problema generalizado se deveu à ausência de um separador entre as seções da receita. Uma busca que ignorasse, por exemplo, todas as ocorrências contendo um “separador” (ou final de linha, ou marca de parágrafo) no meio do padrão recuperado teria evitado várias formas que contêm a seqüência “n + n” e não são UTEs (p. ex.: *sal Modo*, *forno Ingrediente(s)*, *pimenta-do-reino Preparo*, *cm de diâmetro Preparo*, *pitada de sal Modo*). Mais uma vez, um problema que poderia ser resolvido com a etiquetagem das receitas do *corpus*. A essa altura, a melhor maneira de remediar o problema seria acrescentar essas palavras (*Modo*, *Preparo*, *Ingrediente(s)*, etc.) a uma *stoplist* para o substantivo em posição final.

Dentre os 35 padrões iniciados por substantivo pesquisados, há 4 padrões que não recuperaram nenhum *type* com ocorrência maior que 3 no corpus:

- **n + adj + adv + adj + conj + adj** (proveniente de “arroz branco bem lavado e escorrido”): ao observar as 7 ocorrências recuperadas no corpus, vi que havia 1 oc. de “arroz branco bem lavado e escorrido”, mas dentre as outras 6 ocorrências encontrei também 3 ocorrências com um padrão semelhante: “tomates(-caqui) médio(s)/grande(s) bem vermelhos/maduros e firmes”. Uma busca pelo mesmo padrão mas tornando a ocorrência do primeiro adjetivo opcional e forçando a busca por ocorrências somente com a conjunção “e” – **n (+ adj) + adv + adj + “e” + adj** – retornou 76 *types*, mas nenhum com frequência acima de 3. No entanto, ainda que a variedade de ingredientes seja a maior responsável por isso, se o *corpus* fosse lematizado, é possível que um número maior *types* com mais de 3 ocorrências fosse recuperado. Vide, por exemplo, os iniciados por “tomate” com oc. = 1: *tomates-caqui médios bem vermelhos e firmes, tomates bem maduros e firmes, tomates bem vermelhos e firmes, tomates grandes bem vermelhos e firmes, tomates médios bem maduros e firmes, tomates-cereja bem vermelhos e firmes*. Enfim, acredito que o melhor seja eliminar esse padrão das buscas e tentar recuperar essa associação de dois adjetivos da porção final do padrão, “adj + conj + adj”, nas buscas da categoria adjetivo.
- **n + adj + adv + adj + prp + n** (proveniente de “açúcar mascavo bem apertado na xícara”): de fato, nenhum dos *types* recuperados é UTE – esse padrão pode ser eliminado das buscas.
- **n + adj + adv + n** (proveniente de “pistache salgado sem casca”): não me dei conta de que “sem” havia sido etiquetado erroneamente como advérbio. Felizmente, o padrão correto, **n + adj + PRP + n**, já havia sido previsto pela lista de padrões de busca.
- **n + adj + punc + adj + conj + prp + n** (proveniente de “amendoim torrado, salgado e sem pele”): dentre as 10 ocorrências recuperadas, as 2 únicas que poderiam ser consideradas como UTE referem-se a amendoim: uma é “amendoim torrado, salgado e sem pele” e a outra “amendoim salgado, torrado e sem pele” – a estrutura parece estar vinculada à ocorrência de certos adjetivos (*torrad**, *salgad**, *sem pele*, *sem casca*) e uma semente oleaginosa (*amendoim*,

amêndoa, pistache, etc.). Esse tipo de associação semântica pode ser mais bem pesquisada por meio da análise de concordâncias (e não pela busca de padrões morfossintáticos fixos). Uma busca por *torrad** no *corpus* usando o concordanciador do WST confirmou minhas suspeitas – a ordem e a ocorrência dos adjetivos na expressão é variada:

N	Concordance
1	io amargo ralado grosso 2 colheres (chá) de amêndoas sem pele torradas e picadas 1 colher (sopa) de salsinha picada sal e pime
2	de rosquinhas de leite moidas 1/2 xícara de amêndoas sem pele torradas e moidas 4 pêras 1/2 litro de leite 150 g de açúcar 1
3	sca Sal a gosto Para polvilhar 2/3 xícara de amêndoa sem pele torrada e cortada em lascas Preparo Aqueça o forno em temp
4	nfeiteiro 1 colher (sopa) de essência de baunilha 125 g de nozes torradas e moidas Modo de Preparo: Bata por 5 minutos na bate
5	dente de alho socado 1/2 xícara de castanha de caju ligeiramente torrada e picada 300 g de peito de peru assado (sobra ou peito d
6	car 1 vidro de leite de coco (200 ml) 100 g de nozes ligeiramente torradas e moidas 100 g de amêndoas ligeiramente torradas e m
7	ligeiramente torradas e moidas 100 g de amêndoas ligeiramente torradas e moidas Modo de preparo Misture a água c
8	inho simples de vinagre e azeite uns pistaches bem picadinhos, já torrados e salgados . * Adoro um pão quentinho feito em casa e d
9	ara (chá) de óleo de milho 1 xícara (chá) de sementes de girassol torradas e sem sal 2 colheres (sopa) de açúcar mascavo Modo
10	rta 200 g de ervilha fresca 1/3 de xícara de pistache sem casca, torrado e salgado , bem picadinho 1 colher (chá) de açúcar 1/4 d
11	Cajuzinho Ingredientes: 500 g de amendoim se casca torrado e triturado 1 coco ralado médio 2 xícaras (chá) de açúcar
12	das (temperatura ambiente) 1/4 xíc. (chá) de castanha de caju, torrada e salgada 2 pimentas vermelhas em cubinhos miúdos
13	imenta-do-reino moida na hora 3 col. (sopa) de castanha-de- caju torradas e moidas 1 col. (sopa) de manteiga sem sal P
14	doim torrado e moído sem casca 1/4 de xícara castanha de caju torrada e moída 1 xícara de leite de coco 1 colher de chá de ge
15	pa) de açúcar de confeiteiro 1/2 xícara (chá) de castanha de caju torrada e moída Prepare uma calda: leve o açúcar ao fogo mod
16	ra de amendoim torrado e moído 1/2 xícara de castanha de caju torrada e moída 1/4 xícara de óleo 2 colheres (sopa) de azeite-
17	Ingredientes 1 xícara de manteiga 1 xícara de castanha de caju torrada e moída 1/3 xícara de salsa picada 3 colheres (chá) de
18	ha, a cebolinha, 2 colheres (sopa) de azeite de oliva, as nozes, as torradas e o sal. Recheie o peixe com a mistura e amarre-o com
19	s de amaretto 3 claras de ovos extra 3 xícaras (chá) de amêndoas torradas sem pele e moidas 1 e 1/2 xícara (chá) de açúcar cristal
20	silha. Aqueça no microondas por 3 minutos, polvilhe as amêndoas torradas e sirva como acompanhamento. Rendimento:
21	édio por cerca de 40 minutos. Decore com laminas de amêndoas torradas e sirva acompanhado de gomos de limão. O tempero s
22	olher de café de suco de limão 2 colheres de sopa de amêndoas torradas e picadas 3 colheres de sopa de vinho branco sal e pim
23	olher de café de suco de limão 2 colheres de sopa de amêndoas torradas e picadas 3 colheres de sopa de vinho branco sal e pim
24	anteiga 1 envelope de açafraão sal 1/2 xícara (chá) de amêndoas torradas e picadas 2 xícaras de arroz lavado Modo de Preparo:
25	de páprica doce pimentão 8g de tinta de lula 50g de amêndoas torradas e picadas Caldo de peixe (que pode ser feito de cabeç
26	ueno (200 g) torradas de baguete 1 colher de sopa de amêndoas torradas e picadas Modo de Preparo: Coloque o queijo brie em
27	Calorias: 410 por fatia Ingredientes 1 1/2 xícara de amêndoa torrada e moída 1 xícara de farinha de trigo 3/4 xícara de açúcar
28	dientes Massa 1 1/4 xícara de açúcar 1/2 xícara de amêndoa torrada e moída 1 colher (chá) de essência de baunilha 200 g d
29	ubinhos 1/2 xícara de uva passa escura 1/2 xícara de amendoim torrado, salgado e sem pele 1/2 xícara de uma mistura de folhas
30	a de trigo 1 colher (chá) de fermento em pó 150 g de amendoim torrado, descascado e ligeiramente moído Modo de Preparo: Ba
31	milho 2 talos de salsão limpos e em cubos 100 g de amendoim torrado, salgado e sem pele Sal a gosto Molho 1/2 copo de mol

Mesmo entre os padrões que tiveram *types* com ocorrência maior ou igual a 3 no *corpus* há alguns que não recuperaram nenhuma forma que pudesse ser considerada uma UTE:

- **n + adj + adj + adj** (proveniente de “massa folhada laminada congelada”): a característica marcante de todas as formas recuperadas é o fato de um dos três adjetivos da seqüência ser sempre um adjetivo inespecífico (como *pequeno, médio, grande, picado, gelado*). Mas examinei também as linhas de concordância de “massa folhada” no *corpus* etiquetado, usando o concordanciador. Constatei que o problema, que já havia identificado quando da correção da etiquetagem do *corpus* – qual seja, vários adjetivos deverbais terem sido classificados como verbo –, também colaborou para o insucesso da extração, especialmente da forma que deu origem ao padrão, já que tanto

"congelada" quanto "laminada" foram etiquetadas erroneamente como verbo no corpus. O ideal seria rever as etiquetas de verbo atribuídas a palavras terminadas em -ado(s) / -ada(s) / -ido(s) / -ida(s) e, se possível, fazer uma substituição em massa (tomando o devido cuidado para não etiquetar palavras como "melado" ou "amido" também como adjetivo). Por ora, o melhor seria eliminar o padrão das buscas (ainda que tenham sido identificadas na lista das formas recuperadas duas ocorrências da UTE "fermento biológico seco granulado").

- **n + adj + adj + conj + adj** (proveniente de "arroz branco lavado e escorrido"): o mesmo problema observado com o item anterior foi observado aqui: a presença massiva de adjetivos inespecíficos (como *pequeno, médio, grande, picado, gelado, cozido*) – o padrão também poderia ser eliminado das buscas sem grandes perdas para o resultado
- **n + adj + n** (proveniente de "lentilha francesa Puy"): apesar do grande número de ocorrências recuperadas no *corpus* (3.831 no total), esse padrão não retornou, dentre os 241 *types* com mais de 3 oc., nenhuma UTE. Em sua grande maioria, as formas apresentam o mesmo problema da subcategoria "n + n" – a ausência de delimitador entre as seções da receita permitiram a recuperação de várias formas como: *salsinha picada Modo, limão ralada²²⁹ Acessório, preparo antecipado Ingredientes*. O padrão poderia ser eliminado da lista de buscas.
- **n + conj + v + adv** (proveniente de "queijo que derrete bem"): não rendeu nenhuma UTE. Em todas as formas recuperadas, a conjunção é "e". Fiz uma busca no *corpus* etiquetado por "que... bem" e não encontrei nenhuma UTE com frequência maior que 1, ainda que haja 3 oc. de "tampa/vidro/pote **que feche bem**" no corpus:

N	Concordance
1	/n ./punc Em/prp uma/art ca?arola/n que/conj tenha/v tampa/n que/conj feche/v bem/adv ./punc arrume/v os/art ingredientes/
2	va/n (/punc aproximadamente/adv)/punc sal/n Num/prp pote/n que/conj feche/v bem/adv ./punc misture/v a/art azeitona/n pre
3	de/prp queijo/n fontina/n ou/conj de/prp algum/pron outro/pron que/conj derreta/v bem/adv (/punc um/art gruy?re/n ou/conj um
4	r/v um/art ./punc use/v um/art escorredor/n de/prp macarr?o/n que/conj encaixe/n bem/adv em/prp uma/art panela/n)/punc ./
5	pron num/prp pote/n bem/adv grande/adj de/prp vidro/n e/conj que/conj feche/v bem/adv ./punc Numa/prp jarra/n ./punc mistu
6	onj regue/v com/prp o/art creme/n de/prp leite/n ./punc deixe/v que/conj absorva/v bem/adv ./punc acrescente/v a/art tigela/n
7	e/conj polvilhe/v as/art am?ndoas/n ./punc fazendo/v com/prp que/conj fixem/v bem/adv no/prp fundo/adj e/conj nas/pron par

²²⁹ proveniente de "casca de limão ralada"

- **n + prp + v + n** (proveniente de “máquina de moer carne” / “máquina de fazer macarrão”): esse padrão me chamou a atenção pois, dentre todos os que tiveram ocorrência de 0 UTEs nos resultados, esse era o único que se originara de 2 UTEs distintas do glossário da memória de tradução. Fui pesquisar o *corpus* etiquetado para ver o que havia acontecido com a expressão que deu origem ao padrão e observei que a etiquetagem errada de “moer” como substantivo foi um dos motivos do insucesso da extração; o outro foi a omissão de “fazer” em “máquina de fazer macarrão”. De qualquer modo, o melhor seria alterar o padrão de busca “n + prp + n” de modo que recupere também essas ocorrências de “máquina de moer carne” (“n + prp + (v) + n”):

N	Concordance
1	adv bem/adv (/punc pode-se/v tamb?m/adv passar/v pela/prp m?quina/n de/prp moer/n carne/n e/conj depois/adv levar/v a
2	hos/n e/conj passe- os/v ./punc novamente/adv ./punc pela/prp m?quina/n de/prp moer/n carne/n ./punc usando/v a/art parte/
3	odo/n de/prp Preparo/prop ./punc Moa/v o/art lombo/n na/prp m?quina/n de/prp moer/n carne/n ./punc com/prp o/art moed
4	/n at?/prp amaciar/v ./punc Desossar/v e/conj passar/v na/prp m?quina/n de/prp moer/n ./punc Misturar/v a/art carne/n com/
5	e/conj a/art pimenta/n ./punc Passar/v novamente/adv na/prp m?quina/n de/prp moer/n ./punc Levar/v ?/v geladeira/n ./punc
6	t suco/n de/prp lim?o/n e/conj passar/v novamente/adv na/prp m?quina/n de/prp moer/n ./punc Deixar/v o/art trigo/n de/prp
7	peixe/n ./punc Enquanto/adv isso/pron ./punc passe/v na/prp m?quina/n de/prp moer/n ./punc em/prp rolo/n fino/adj ./punc
8	prop muito/adv bem/adv ./punc Passe/v a/art mistura/n na/prp m?quina/n de/prp moer/n ./punc com/prp a/art pe?a/n grossa
9	por/prp 30/num minutos/n ./punc Abra/v a/art massa/n na/prp m?quina/n de/prp macarr?o/n e/conj polvilhe-a/n com/prp fari
10	ssa/n em/prp tiras/n ./punc com/prp o/art rolo/n ou/conj na/prp m?quina/n de/prp macarr?o/n ./punc e/conj corte-a/v em/prp
11	utos/n ./punc A/art seguir/v ./punc abra/v a/art massa/n na/prp m?quina/n de/prp macarr?o/n (/punc ou/conj com/prp um/art
12	ada/n de/prp noz-moscada/n Sal/n a/art gosto/n Acess?rios/n M?quina/n de/prp macarr?o/n Cortador/n de/prp massa/n Pr
13	dv ./punc abra/v a/art massa/n bem/adv fino/adj com/prp a/art m?quina/n de/prp macarr?o/n e/conj corte/v tiras/n de/prp 10,

De maneira geral, os padrões iniciados por substantivo que tiveram melhor desempenho na extração automática de UTEs foram (vide mais exemplos no Anexo V):

- **n + adj + conj + adj** (ex.: arroz lavado e escorrido)
- **n + adj + prp + n + adj** (ex.: fôrma redonda de aro removível)
- **n + adv + adj** (ex.: assadeira não untada)
- **n + n + adj** (ex.: lingüiça calabresa defumada)
- **n + n + punc + n + prp + n + punc** (o ideal seria escrever uma expressão regular para ampliar a busca desse padrão, que pegou apenas “nam pla (molho de peixe)”, mas poderia abranger também todas as ocorrências que incluem os parênteses)

B. Padrões iniciados por preposição

A pesquisa inicial foi feita considerando expressões iniciadas por quaisquer preposições, o que totalizou 234.345 oc. no corpus. No entanto, um exame rápido dos resultados mostrou que poderia ser vantajoso eliminar todas as ocorrências iniciadas por: "de", "do", "dos", "da", "das", "De", "Do", "Dos", "Da", "Das", "DE", "DO", "DA", "DOS", "DAS", "para", "Para", "PARA", "SOBRE", "sobre" e "Sobre", ainda que expressões como “de véspera” e “de vez em quando” fossem perdidas. Escrevi uma rotina em Python que fazia a busca dos padrões ignorando os iniciados por essas preposições. O novo resultado foi usado para a análise randômica das formas recuperadas.

O número total de ocorrências diminuiu drasticamente (de 234.343 para 68.072), o número total de *types* (23.184) e o número de *types* com oc. maior ou igual 3 (8.003) caíram, respectivamente, para 13.022 e 3.502. Talvez isso justifique a acurácia de 25% obtida na análise da amostra aleatória (teria sido muito menor, certamente, se tivesse sido analisada uma amostra randômica do resultado que incluísse quaisquer preposições na posição inicial). Entretanto, para me certificar de que não havia perdido nenhum *type* muito importante iniciado pelas preposições da *stoplist*, foi feita uma análise *en passant* de todas as formas obtidas sem a *stoplist*. É com base nessa análise que teço os comentários a seguir sobre essa categoria.

De modo geral, observei que seria melhor fazer uma *stoplist* específica para os padrões que renderam muitos *types*, mas poucas UTEs. Alguns, como “prp + v” (*em seguida, para decorar*) e “prp + n + adj” (*no dia seguinte, em rodela grossas*) por exemplo, já se beneficiariam com a eliminação apenas de formas iniciadas por “ao” e “com”; já no caso de “prp + n + adv + adj” (*em rodela bem finas*), a preposição “de” teria que fazer parte dessa lista também. No caso de “prp + art + n + prp + v” (*até a hora de servir*), por outro lado, o ideal seria ignorar *types* em que a segunda preposição fosse igual a “até” – que, por sua vez, é a preposição em posição inicial que rendeu mais UTEs não só nesse padrão mas em “prp + v + adj” (*até que esteja al dente*), “prp + v + prp + n” (*até dobrar de volume*) e “prp + conj + v + adj” (*até que fiquem macias*).

Dentre os padrões que não recuperaram nenhum *type* com oc. maior ou igual a 3 observei o seguinte:

- **prp + art + n + adj + prp + adj** (proveniente de “com a boca virada para baixo”): aqui, mais uma vez, a etiquetagem errônea dos adjetivos deverbais como verbo

foi o que prejudicou a extração dos padrões. Uma busca no *corpus* pelo padrão `com/\S+\s+\S+/art\s+\S+/n\s+[\S+/\S+\s+]{0,20}?\para/\S+\s+\S+/\S+\s+` (“com” + art + n + ... + “para” + ...) retornou 366 ocorrências, sendo 271 types, dos quais 21 tiveram oc. maior ou igual a 3 no corpus. Um exame dos 34 types com 2 ocorrências, a exemplo do que acontecera em outros padrões de busca maiores, revelou um grande número de UTEs (11 – 32% do total) que poderiam ser agrupadas sob um mesmo padrão (“com” + **ART DEF** + “parte / lado” / **PARTE DO ALIMENTO/RECIPIENTE** (+ “voltada / virada”) + “para” + “cima / baixo”), juntamente com as UTEs com oc. ≥ 3:

UTES	forma recuperada	oc. no corpus
1	com/prp a/art pele/n voltada/v para/prp baixo/adj	10
2	com/prp a/art pele/n para/prp baixo/adj	8
3	com/prp a/art parte/n cortada/v para/prp baixo/adj	4
4	com/prp a/art pele/n para/prp cima/n	3
5	com/prp a/art pele/n virada/v para/prp baixo/adj	3
6	com/prp o/art peito/n virado/v para/prp baixo/adj	3
7	com/prp a/art parte/n cortada/v para/prp cima/n	3
8	com/prp o/art lado/n da/prp papoula/n para/prp baixo/adj	2
9	com/prp o/art lado/n da/prp crosta/n para/prp baixo/adj	2
10	com/prp o/art peito/n para/prp cima/n	2
11	com/prp a/art polpa/n voltada/v para/prp cima/n	2
12	com/prp o/art lado/n amanteigado/v para/prp baixo/adj	2
13	com/prp o/art lado/n cortado/adj para/prp cima/n	2
14	com/prp o/art lado/n da/prp gordura/n para/prp cima/n	2
15	com/prp o/art peito/n virado/v para/prp cima/n	2
16	com/prp a/art parte/n da/prp casca/n para/prp baixo/adj	2
17	com/prp a/art boca/n para/prp cima/n	2
18	com/prp o/art lado/n da/prp pele/n para/prp cima/n	2
total		56

- **prp + conj + v + adj + conj + adj** (proveniente de “até que esteja macio e dourado”): o grande problema aqui foi a variação nos adjetivos que compõem o subpadrão “adj + conj + adj”, especialmente quanto à combinatória de itens lexicais (*macio e firme, firmes e secos, mole e macio, opaca e macia*). No entanto, houve 3 ocorrências de “lisa e homogênea” (uma começando com “para que fique” e 2 com “até que fique”) e uma oc. de “até que fique lisa e elástica”. Um consulta ao *corpus* usando o concordanciador revela que há um padrão longo e com porções opcionais e variáveis que poderia ser descrito como

“VERBO até (+que) + v + ... + adj (+ conj + adj)”, sendo que “adj (+ conj + adj)” indicam um qualificativo da aparência e/ou consistência desejada do alimento (*até que estejam macios, até formar um creme liso, até que fiquem dourados e crocantes, até obter uma massa lisa e elástica, até obter uma mistura homogênea*, etc.). O melhor seria, a meu ver, excluir este padrão da lista e fazer buscas específicas com “até” usando o concordanciador.

- **prp + prp + adv + num + n + prp + n** (proveniente de “com pelo menos X horas de antecedência”): o fato de os números terem sido mantidos no *corpus* (em vez de serem transformados em uma variável, do tipo “X” – o que é geralmente feito ao etiquetar um corpus) foi o motivo principal do insucesso da coleta, já que 5 das 7 formas recuperadas são exemplos idênticos ao que originou o padrão, mudando apenas o numeral usado. Ainda assim, uma busca no *corpus* por “antecedência” mostra que uma expressão de busca mais abrangente seria capaz de identificar outras variações desse padrão, como: *com umas X horas de antecedência, com até X dias de antecedência, com no máximo X horas de antecedência*. Este seria um padrão possível para fazer essa busca: `com/prp\s+[\S+\s+]{0,60}?\S+/num\s+\S+/n\s+\S+/prp\s+anteced\S+\S+\s+`
- **prp + v + conj + v + art + n + prp + n** (proveniente de “até encorpar e cobrir o dorso da colher”): a variação, mais uma vez, impediu que o padrão recuperasse os diversos *types* que podem representar a UTE que originou a busca. Dentre os 15 recuperados, 2 representam a UTE original, um começando com “até engrossar” e outro com “até encorpar”. Uma busca no *corpus* por “cobrir” e “colher” usando o concordanciador revela que “até” pode estar bem distante de “cobrir”:

N	Concordance
1	/punc o/art suco/n de/prp lim?o/n e/conj 1/num pitada/n de/prp sal/n (/punc se/conj o/art creme/n for/v fresco/adj ,/punc bata/v at?/prp encorpar/v e/conj cobrir/v o/art dorso/n da/prp colher/n e/conj depois/adv junte/vo/art lim?o/n e/conj o/art sal/n)/punc ./punc Passe/v para/prp uma/art tigela/n ,/pu
2	ovamente/adv ./punc Continue/v a/art mexer/v sem/prp parar/v ,/punc at?/prp sumir/v a/art espuma/n da/prp superf?cie/n ,/punc engrossar/v e/conj cobrir/v o/art dorso/n da/prp colher/n ,/punc mas/conj sem/prp ferver/v ./punc Imediatamente/adv ,/punc passe/v o/art creme/n por/prp uma/art p
3	/conj ,/punc sem/prp parar/v de/prp mexer/v ,/punc mantenha/v no/prp fogo/n at?/prp a/art espuma/n sumir/v ,/punc o/art creme/n engrossar/v e/conj cobrir/v o/art dorso/n da/prp colher/n ,/punc Para/prp que/conj o/art creme/n n?o/adv ferva/v nem/conj talhe/v ,/punc transfira/v imediatamente/a
4	e/n de/prp leite/n e/conj ,/punc sempre/adv mexendo/v ,/punc deixe/v ferver/v por/prp uns/pron 5/num minutos/n ,/punc at?/prp engrossar/v e/conj cobrir/v o/art dorso/n da/prp colher/n (/punc se/conj quiser/v ,/punc prepare/v na/prp v?spera/n e/conj deixe/v na/prp geladeira/n)/punc ./punc D
5	/n ./punc Misture/v o/art vinagre/n e/conj o/art a??car/n numa/prp panela/n m?dia/adj ,/punc aque?a/v e/conj deixe/v ferver/v at?/prp encorpar/v e/conj cobrir/v o/art dorso/n da/prp colher/n ,/punc depois/adv junte/vo/art caldo/n reservado/adj e/conj deixe/v no/prp fogo/n at?/prp encorpar/v de/pr
6	icione/v tamb?m/adv o/art creme/n de/prp leite/n e/conj ,/punc sem/prp parar/v de/prp mexer/v ,/punc deixe/v no/prp fogo/n at?/prp engrossar/v e/conj cobrir/v o/art dorso/n da/prp colher/n ,/punc Junte/v a/art carne/n com/prp a/art cenoura/n ,/punc a/art cebolinha/n e/conj o/art cogumelo/n ,/punc acerte/
7	t?/prp surgirem/v bolinhas/n nas/pron laterais/adj ,/punc sumir/v a/art espuma/n da/prp superf?cie/n ,/punc o/art creme/n encorpar/v e/conj cobrir/v o/art dorso/n da/prp colher/n ,/punc mas/conj sem/prp ferver/v ./punc Despeje/prop metade/n do/prp creme/n sobre/prp o/art chocolate/
8	la/n e/conj ,/punc sem/prp parar/v de/prp mexer/v ,/punc deixe/v no/prp fogo/n at?/prp a/art espuma/n sumir/v e/conj o/art creme/n engrossar/v e/conj cobrir/v o/art dorso/n da/prp colher/n ,/punc Despeje/prop imediatamente/adv sobre/prp o/art chocolate/n e/conj misture/v at?/prp derreter/v ./p

Mesmo entre os padrões que tiveram *types* com ocorrência maior ou igual a 3 no *corpus* há alguns que não recuperaram nenhuma (ou quase nenhuma) forma no *corpus* que pudesse ser considerada uma UTE:

- **prp + art + n + adv** (proveniente de “em um pedaço só”): primeiramente fiz uma busca no *corpus* para ver as ocorrências de “só” à direita e à esquerda de “pedaço” – surpreendentemente, as 15 ocorrências são de “em um só pedaço” (sinal de que eu não havia pesquisado a Internet ou o *corpus* ao acrescentar essa UTE no glossário de minha memória de tradução!). Observando os resultados obtidos (p. ex.: *com uma faca bem*, *com a panela parcialmente*, *com o óleo bem*), constatei que se esse padrão fosse reescrito de duas formas diferentes seria muito mais eficiente na extração: “**prp + art + n + adv + adj**” (para recuperar *com uma faca bem afiada*, *com a panela parcialmente tampada*, p. ex.) e “**prp + art + n + adv + v (+ adj)**” (para recuperar *para o caldo não secar*, *para a massa não ficar pesada*, p. ex.).
- **prp + art + n + prp + n + adj** (proveniente de “com a batedeira em velocidade alta”): todos os *types* com oc. ≥ 3 recuperados não são UTEs. Foi buscado no *corpus*, então, a expressão que originou o padrão digitando “batedeira” e

“velocidade” nas janelas de busca. Há 22 ocorrências no *corpus* dessa UTE, com ligeiras variações (a mais insignificante delas – “com” e “Com” foi responsável pelo insucesso da extração dos padrões idênticos ao original, pois dissociou as ocorrências dessa UTE iniciadas por letras maiúsculas e minúsculas). Pelas concordâncias é possível observar também que o padrão poderia ser reescrito para recuperar 20 dessas ocorrências: **prp (+ art) + n (...) + prp + n + adj**.

N	Concordance
1	/v ./punc Junte/v o/art a??car/n ./punc Bata/v com/prp a/art batedeira/n em/prp alta/adj velocidade/n por/prp 8/num min
2	onj 1/2/num x?caras/n de/prp farinha/n ./punc Com/prp a/art batedeira/n em/prp baixa/adj velocidade/n ./punc adicione/
3	/n ./punc Bata/v os/art ovos/n com/prp o/art a??car/n na/prp batedeira/n em/prp velocidade/n alta/adj ./punc at?/prp obt
4	p 90/num ?C/n ./punc Coloque/v essa/pron mistura/n na/prp batedeira/n em/prp velocidade/n alta/adj at?/prp esfriar/v ./
5	inha/n de/prp trigo/n e/conj fermento/n ./punc Com/prp a/art batedeira/n em/prp velocidade/n alta/adj bata/v a/art mante
6	rt creme/n de/prp leite/n na/prp mesma/pron tigela/n da/prp batedeira/n em/prp velocidade/n alta/adj por/prp 3/num min
7	o/art chocolate/n numa/prp tigelinha/n ./punc Com/prp a/art batedeira/n em/prp velocidade/n baixa/adj ./punc bata/v a/
8	tigela/n limpa/adj e/conj seca/n ./punc bata/v com/prp a/art batedeira/n em/prp velocidade/n baixa/adj at?/prp espuma
9	/n numa/prp tigela/n grande/adj e/conj bata/v com/prp a/art batedeira/n em/prp velocidade/n baixa/adj at?/prp surgir/v
10	loque/v em/prp outro/pron recipiente/n ./punc Com/prp a/art batedeira/n em/prp velocidade/n m?nima/adj v?/v acresce
11	s/pron para/prp incorporar/v bem/adv ./punc Com/prp a/art batedeira/n ligada/v em/prp velocidade/n baixa/adj despej
12	s/pron para/prp incorporar/v bem/adv ./punc Com/prp a/art batedeira/n ligada/v em/prp velocidade/n baixa/adj despej
13	ela/n a/art farinha/n e/conj o/art fub?/n ./punc Com/prp a/art batedeira/n ligada/v em/prp velocidade/n baixa/adj ./punc v
14	nc Bata/v as/art claras/n em/prp neve/n e/conj com/prp a/art batedeira/n ligada/v em/prp velocidade/n baixa/adj adicion
15	o/art vinagre/n ?/v ?gua/n fervente/adj ./punc Com/prp a/art batedeira/n ligada/v em/prp velocidade/n m?dia/adj v?/v ac
16	sal/n ./punc Derreta/v a/art manteiga/n ./punc Com/prp a/art batedeira/n ligada/v em/prp velocidade/n m?nima/adj ./pun
17	n)/punc de/prp fermento/n em/prp p?/n ./punc Bata/v na/prp batedeira/n ./punc em/prp velocidade/n alta/adj ./punc 4/nu
18	./punc Bata/v o/art creme/n de/prp leite/n fresco/adj na/prp batedeira/n ./punc em/prp velocidade/n alta/adj ./punc por/
19	a/prp ?gua/n que/conj foi/v congelada/v e/conj bata/v na/prp batedeira/n ./punc em/prp velocidade/n alta/adj ./punc por/
20	te/n em/prp p?/n aos/prp poucos/pron e/conj com/prp a/art batedeira/n ./punc em/prp velocidade/n m?nima/adj ./punc
21	eve/n bem/adv firme/n e/conj reserve/v ./punc Bata/v na/prp batedeira/n ./punc em/prp velocidade/n m?nima/adj (/punc
22	r/v homog?neo/adj ./punc A/art seguir/v ./punc bata/v na/prp batedeira/n ./punc em/prp velocidade/n m?xima/adj ./punc

Por fim, chama a atenção a regularidade e a alta frequência das UTEs iniciadas por preposição – o que endossa a hipótese de que esses padrões não podem ser desconsiderados na compilação de dicionários técnicos voltados para o tradutor. Vide, por exemplo, a precisão e o alto número de UTEs recuperadas pelos seguintes padrões (vide mais exemplos no Anexo V):

- **prp + v + prp + v:** (ex.: sem parar de mexer) 6 *types* que renderam 650 *tokens*, todas UTEs;
- **prp + v + adj:** (ex.: até ficarem macias) 33 *types* que renderam 794 *tokens*, a maioria é UTE e começa com “até”;
- **prp + conj + v + prp + n:** (ex.: até que dobre de volume) 4 *types*, 29 *tokens*, todas são UTEs

- **prp + conj + v + art + n (+...) + pron + v + adj**: esse padrão, transformado numa expressão regular recuperou 78 ocorrências de uma única UTE, uma das mais longas dentre as pesquisadas: “até que enfiando um palito (no ALIMENTO ASSADO) ele saia limpo” (há ainda outras variações, como “enfiando / espetando” um(a) “garfo / faca”);
- **prp + art + adv + n**: (ex.: em uma só camada) 4 *types* que geraram 135 *tokens*.

C. Padrões iniciados por verbo

Os padrões iniciados por verbo foram os que tiveram o maior nível de precisão na extração de *types* considerados como UTEs – 50% da amostra aleatória. De forma geral, um dos maiores problemas encontrados para esses padrões foi o fato de o verbo de ligação “é” ter sido etiquetado da mesma forma que a preposição “à” (já que ambos transformaram-se num sinal de interrogação por serem caracteres acentuados). No entanto, observou-se que o verbo “é” na posição inicial do padrão não rendeu UTEs, de forma que uma *stoplist* eliminando todas essas ocorrências resolveria os dois problemas de uma só vez. Outra *stoplist* útil para eliminar ruídos em alguns padrões seria com os verbos de suporte (esteja(m), fique(m), espere, vá, deixe) na posição final do padrão.

Os padrões que não recuperaram nenhum *type* com ocorrência ≥ 3 foram analisados e observou-se que:

- **v + art + n + adj + prp + adj** (proveniente de *junte a calda quente aos poucos*): é evidente a variação possível numa UTE como esta, que seria mais produtiva se transformada numa expressão regular do tipo “v (+ n) aos poucos/poquinhos” (que recupera nada menos que 958 ocorrências da UTE no corpus). E, como já vimos ao discutir os padrões iniciados por preposição, “poucos” foi etiquetado como pronome no corpus, e “pouquinhos” como substantivo, de forma que a expressão, como está, pode ser eliminada da lista de padrões.
- **v + art + n + adj + prp + pron** (proveniente de: *deixando um espaço livre entre elas*): ao analisar os 40 *types* com oc. menor que 3 no *corpus* recuperados por esse padrão observei que havia 5 ocorrências de “mantendo/deixando um espaço livre entre eles/elas”. Foi feita uma busca de “espaço livre” no *corpus* usando o concordanciador e encontradas mais duas ocorrências: uma com um

advérbio antes de “espaço”, e outra omitindo o artigo. Uma expressão regular que daria conta dessas variações, mas não valeria a pena usá-la, já que o caso aqui é muito mais de associação semântica ao item lexical “espaço livre” do que de estrutura morfossintática:

N	Concordance
1	/prp contato/n com/prp o/art caramelo/n ./punc sem/prp deixar/v espaço/n livre/adj ./punc Reserve/v ./punc Bata/v as/art claras/n
2	de/prp massa/n nas/pron assadeiras/n ./punc mantendo/v um/art espaço/n livre/adj de/prp uns/pron 2/num cm/n entre/prp eles/pr
3	um cm/n ./punc distribua/v na/prp assadeira/n deixando/v um/art espaço/n livre/adj de/prp uns/pron 2/num cm/n entre/prp elas/pr
4	um cm/n e/conj espalhe/v na/prp assadeira/n mantendo/v um/art espaço/n livre/adj entre/prp eles/pron ./punc Cubra/v os/art p?ez
5	4/num de/prp x?cara/n de/prp massa/n ./punc mantendo/v um/art espaço/n livre/adj entre/prp elas/pron (/punc a/art massa/n ?/v s
6	olheres/n de/prp ch?/n de/prp massa/n ./punc deixando/v um/art espaço/n livre/adj entre/prp elas/pron ./punc espere/v dourar/v n
7	la/prp espalhe/v os/art bifés/n ./punc mantendo/v um/art bom/adj espaço/n livre/adj entre/prp eles/pron ./punc Doure/v bem/adv d

Quanto a padrões que recuperaram *types* com oc. ≥ 3 mas nenhum deles constituindo UTE, foram encontrados os seguintes:

- **v + art + n + adj + prp + art + n** (proveniente de *deixando um espaço livre entre os pães*): estrutura muito semelhante a “deixando um espaço livre entre elas”, discutida logo acima;
- **v + art + n + prp + n + adv** (proveniente de *leve a panela ao fogo novamente e leve a frigideira ao fogo novamente*): aqui observei um fato muito interessante. Se for feita uma busca no *corpus* por “fogo/forno” e “novamente” ocorrendo à direita, há 30 ocorrências, sendo que “leve ao fogo novamente” ocorre 22 vezes, e “leve ao forno novamente” 6 vezes (“leve a frigideira ao fogo novamente” e “coloque-a [panela] no fogo novamente” ocorrem apenas 1 vez cada). No entanto, se a busca com “novamente” ocorrendo apenas à esquerda é repetida, há mais variação: entre as 70 ocorrências, o verbo “coloque” ocorre 6 vezes, e em 8 contextos há um ingrediente (“bolo”, “frango”) e não um utensílio, como “frigideira”) entre o verbo e “ao fogo / forno” – os dois casos em que ocorre um utensílio são com o verbo “coloque”;
- **v + art + n + conj + v + prp + art + n** (proveniente de *retire a espuma que subir com a escumadeira*): a variação lexical e de posição das palavras na frase, como pode ser visto nas linhas de concordâncias geradas para “espuma” e “subir” no *corpus*, foram as grandes responsáveis pelo insucesso da extração:

N	Concordance
1	/prp uma/art escumadeira/n ./punc retire/v cuidadosamente/adv a/art espuma/n formada/v na/prp superf?cie/n ./punc Em/prp seguida/v ./p
2	m/prp que/conj ir?/v preparar/v as/art crepes/n ./punc descarte/v a/art espuma/n que/conj subir/v ./punc retire/v do/prp fogo/n assim/adv qu
3	ros/adj e/conj a/art pimenta/n e/conj aque?a/v ./punc Descarte/v a/art espuma/n que/conj subir/v ./punc abaixe/v o/art fogo/n quando/adv fe
4	punc cubra/v com/prp ?gua/n e/conj aque?a/v ./punc Descarte/v a/art espuma/n que/conj subir/v e/conj deixo/v no/prp fogo/n at?/prp que/c
5	j o/art couro/n de/prp bacon/n ./punc Aque?a/v ./punc descarte/v a/art espuma/n que/conj subir/v e/conj cozinhe/v em/prp fogo/n baixo/adj a
6	/n m?dia/adj ./punc Derreta/v a/art manteiga/n ./punc descarte/v a/art espuma/n que/conj subir/v e/conj deixo/v amornar/v ./punc Divida/v o/
7	dado/v no/prp fundo/adj ./punc Deixe/v ferver/v ./punc descarte/v a/art espuma/n que/conj subir/v com/prp uma/art escumadeira/n ./punc vol
8	ue?a/v a/art manteiga/n numa/prp panelinha/n ./punc descarte/v a/art espuma/n que/conj subir/v com/prp uma/art escumadeira/n ./punc tra
9	c sal/n e/conj pimenta/n ./punc deixo/v ferver/v ./punc descarte/v a/art espuma/n que/conj subir/v ./punc abaixe/v o/art fogo/n e/conj cozinhe
10	t manteiga/n em/prp fogo/n bem/adv baixo/adj e/conj descarte/v a/art espuma/n que/conj fica/v na/prp superf?cie/n ./punc Deixe/v descans
11	/v a/art manteiga/n numa/prp panelinha/n e/conj descarte/v toda/pron espuma/n que/conj subir/v com/prp uma/art escumadeira/n ./punc Es
12	nj misture/v bem/adv ./punc Deixe/v ferver/v ./punc eliminando/v a/art espuma/n que/conj se/conj formar/v na/prp superf?cie/n durante/prp
13	nj com/prp a/art ajuda/n de/prp uma/art escumadeira/n elimine/v a/art espuma/n que/conj se/conj forma/n na/prp superf?cie/n ./punc Tamp
14	0/num minutos/n ./punc retire/v com/prp uma/art escumadeira/n a/art espuma/n que/conj se/conj forma/n na/prp superf?cie/n ./punc Ap?s/
15	o/art cozimento/n ./punc retire/v com/prp uma/art escumadeira/n a/art espuma/n que/conj formar/v na/prp superf?cie/n ./punc Retire/v do/pr
16	ando/adv ferver/v ./punc retire/v com/prp uma/art escumadeira/n a/art espuma/n que/conj subir/v ./punc abaixe/v o/art fogo/n ./punc junte/v
17	retirando/v com/prp a/art ajuda/n de/prp uma/art escumadeira/n a/art espuma/n que/conj se/conj forma/n na/prp superf?cie/n ./punc Passe
18	/n ./punc Seja/prop no/prp contraste/n das/prp cores/n ou/conj na/prp espuma/n que/conj convida/v ao/prp consumo/n imediato/adj ./punc
19	num por??es/n Calorias/n ./punc 110/num Cal/prop por/prp por??o/n Espuma/prop rosada/v Ingredientes/n ./punc 1/num x?cara/n (/punc c
20	o/n baixo/adj ./punc retirando/v de/prp vez/n em/prp quando/adv a/art espuma/n que/conj se/conj formar/v na/prp superf?cie/n ./punc Coloq
21	/n e/conj ./punc com/prp uma/art escumadeira/n ./punc remova/v a/art espuma/n que/conj se/conj formar/v na/prp superf?cie/n ./punc Quan
22	eixe/v o/art doce/adj ferver/v em/prp fogo/n baixo/adj retirando/v a/art espuma/n que/conj se/conj forma/n na/prp superf?cie/n e/conj mistur
23	/n ./punc Ferver/v por/prp 30/num min/n ./punc ./punc retirando/v a/art espuma/n que/conj se/conj forma/n na/prp superf?cie/n ./punc Apaga
24	om/prp a/art ?gua/n ./punc Leve/v ?/v fervura/n ./punc retirando/v a/art espuma/n que/conj se/conj forma/n na/prp superf?cie/n ./punc Ferva/
25	rt cenoura/n e/conj ferva/v por/prp 20/num minutos/n retirando/v a/art espuma/n que/conj se/conj forma/n na/prp superf?cie/n ./punc Acres
26	om/prp a/art ?gua/n ./punc Leve/v ?/v fervura/n ./punc retirando/v a/art espuma/n que/conj se/conj forma/n na/prp superf?cie/n ./punc Ferva/
27	punc a/art pimenta/n por/prp 10/num minutos/n ./punc retirando/v a/art espuma/n que/conj se/conj formar/v ./punc Em/prp seguida/v ./punc j
28	nj deixo/v cozinhar/v por/prp 2/num minutos/n ./punc retirando/v as/art espuma/n que/conj se/conj formar/v na/prp superf?cie/n ./punc Coe/v
29	ter/v a/art manteiga/n ./punc Retire/v do/prp fogo/n e/conj retire/v a/art espuma/n da/prp superf?cie/n ./punc Junte/v 1/num colher/n (/punc s
30	limpa/adj ./punc aque?a/v ./punc espere/v ferver/v ./punc retire/v a/art espuma/n e/conj cozinhe/v em/prp fogo/n baixo/adj at?/prp que/conj
31	to/n ./punc Deixe/v ferver/v em/prp fogo/n alto/adj ./punc Retire/v a/art espuma/n e/conj adicione/v 4/num talos/n de/prp sals?o/n ./punc 1/nu

- **v + conj + v + prp + n** (proveniente de *unte e polvilhe com farinha*): o maior problema com esse padrão é que, na verdade, ele tem por base a associação lexical “unte e polvilhe (com farinha)”, e não uma estrutura morfossintática. A propósito, essa associação ocorre muito mais vezes na forma de sintagma adjetival “untada e polvilhada (com farinha)” no corpus. O padrão pode, portanto, ser eliminado da lista sem nenhuma perda para a extração automática
- **v + v + prp + adj** (proveniente de *vá juntando aos poucos*): a variação, mais uma vez, foi a responsável pela ineficiência da coleta. Uma busca no *corpus* por (“vá” “aos poucos / pouquinhos”) retornou 66 ocorrências: 25 com “acrescentando”, 11 com “adicionando”, 8 com “colocando” e 5 com “juntando” (as demais ocorreram apenas 1 vez cada). Os ingredientes presentes na expressão aparecem com mais frequência entre o verbo e a expressão (*vá juntando a água aos poucos*), mas também depois (*vá colocando aos poucos a farinha*). Seria possível, portanto, alterar o padrão de busca para abrigar essas variações e mantê-lo na lista.

Com relação aos padrões mais produtivos, os que representam estruturas maiores são os que tiveram maior precisão, ainda que nem todas as ocorrências da UTE e suas variações tenham sido recuperadas. Os padrão “v + art + n + art + num +

art” (proveniente de *junte os ovos um a um*), por exemplo, recuperou 3 *types* que são variações dessa UTE, com um total de 10 *tokens*. Mas uma busca no *corpus* por “um a um / uma a uma” revela um total de 138 oc. (a grande maioria, é verdade, ocorrendo com “ovos / gemas”). Um dos problemas do baixo desempenho desse padrão foi a etiquetagem errada de “um / uma” ora como artigo, ora como pronome (em vez de numeral). O padrão “v + art + n + pron + art + art” (*quebre os ovos um a um*) foi criado justamente para tentar minimizar esse erro, mas o ideal seria juntar os dois numa única expressão regular.

Há também outras variantes menos recorrentes dessa UTE que não são recuperadas por esses dois padrões, como as que podem ser vistas nas linhas de concordância a seguir, cujas palavras de busca foram “um*” e “vez” à direita do nódulo (38 oc. obtidas no total):

N	Concordance
1	uto/n antes/adv de/prp retirar/v da/prp assadeira/n ./punc Asse/v uma/art assadeira/n por/prp vez/n no/prp centro/n do/prp forno/n
2	uto/n antes/adv de/prp retirar/v da/prp assadeira/n ./punc Asse/v uma/art assadeira/n por/prp vez/n no/prp centro/n do/prp forno/n
3	ereter/v ./punc Acrescente/v a/art farinha/n de/prp trigo/n ./punc uma/art colher/n por/prp vez/n ./punc misturando/v a/art cada/pr
4	prp parar/v de/prp bater/v ./punc adicione/v o/art a??car/n (./punc uma/art colher/n por/prp vez/n)/punc ./punc Continue/v a/art bat
5	t pitada/n de/prp sal/n ./punc Acrescente/v as/art gemas/n ./punc uma/art de/prp cada/pron vez/n ./punc batendo/v sempre/adv ./p
6	/art creme/n claro/adj ./punc Acrescente/v as/art gemas/n ./punc uma/art de/prp cada/pron vez/n ./punc batendo/v sempre/adv ./p
7	/n cremosa/adj ./punc 2/num ./punc Junte/v as/art gemas/n ./punc uma/art de/prp cada/pron vez/n ./punc o/art coco/n ./punc o/art c
8	n e/conj a/art noz-moscada/n ./punc Junte/v as/art claras/n ./punc uma/art de/prp cada/pron vez/n ./punc misturando/v bem/adv ./p
9	?a/v o/art forno/n ?/v temperatura/n m?dia/adj ./punc Retirando/v um/art pacote/n por/prp vez/n da/prp geladeira/n ./punc modele/
10	ata/v por/prp 5/num minutos/n ./punc Junte/v os/art ovos/n ./punc um/art por/prp vez/n ./punc batendo/v bem/adv a/art cada/pron a
11	/2/num colher/n sopa/n)/punc ./punc Junte/v os/art ovos/n ./punc um/art por/prp vez/n ./punc o/art sal/n ./punc a/art farinha/n de/pr
12	/adv ./punc Adicione/v os/art ingredientes/n restantes/adj ./punc um/art por/prp vez/n ./punc batendo/v sempre/adv ./punc at?/prp
13	p Grand/prop Marnier/prop ./punc Passe/v os/art crepes/n ./punc um/art por/prp vez/n ./punc na/prp mistura/n de/prp licor/n ./punc
14	e/prp mexer/v ./punc Adicione/v ?/v massa/n os/art ovos/n ./punc um/art por/prp vez/n ./punc a/art couve-flor/n ./punc o/art queijo/n
15	rp obter/v picos/n firmes/adj ./punc junte/v as/art gemas/n ./punc uma/art por/prp vez/n ./punc batendo/v sempre/adv at?/prp a/art
16	ente/adj ./punc Coloque/v as/art por??es/n de/prp alfafa/n ./punc uma/art por/prp vez/n ./punc e/conj frite/v at?/prp dourar/v (./punc
17	/n cremoso/adj com/prp as/art gemas/n ./punc adicionando- as/v uma/art por/prp vez/n ./punc Bata/v as/art claras/n em/prp ponto/
18	/art panela/n n?o/adv for/v muito/adv grande/adj ./punc cozinhe/v um/art por/prp vez/n ./punc Creme/n ./punc derreta/v a/art mante
19	c Incorpore/v o/art chocolate/n derretido/v 2/num gemas/n ./punc uma/art por/prp vez/n ./punc Adicione/v as/art claras/n e/conj o/a
20	ferver/v ./punc baixe/v o/art fogo/n e/conj v?/v colocando/v ./punc um/art por/prp vez/n e/conj com/prp bastante/adv cuidado/n ./pu
21	num/prp palito/n pr?prio/adj para/prp churrasco/n e/conj banhe/v uma/num de/prp cada/pron vez/n no/prp chocolate/n derretido/v
22	te/adv ./punc Reduza/v o/art fogo/n e/conj junte/v os/art crepes/n um/num de/prp cada/pron vez/n ./punc Deixe/v cozinhar/v por/pr
23	m/prp parar/v de/prp bater/v ./punc acrescente/v as/art gemas/n uma/num de/prp cada/pron vez/n ./punc Bata/v por/prp mais/adv
24	eiramente/adv em/prp papel/n absorvente/adj e/conj passe-os/v um/num de/prp cada/pron vez/n pela/prp massa/n ./punc Depois
25	ma/art mistura/n cremosa/adj ./punc Junte/v os/art ovos/n ./punc um/num de/prp cada/pron vez/n ./punc e/conj bata/v bem/adv ./p
26	ra/n cremosa/adj ./punc 2/num ./punc Junte/v os/art ovos/n ./punc um/num de/prp cada/pron vez/n ./punc batendo/v sempre/adv ./p
27	bter/v um/art creme/n fofo/adj ./punc Junte/v os/art ovos/n ./punc um/num de/prp cada/pron vez/n ./punc batendo/v at?/prp mistur

Os padrões iniciados por “v + prp + n” renderam todos muitas UTEs, ainda que alguns *types* de alta recorrência tenham que ser transformados em padrões semânticos, como:

- leve a(o) FONTE DE CALOR/FRIO por (pelo menos / no máximo/mínimo / um período de) NUM UNID. DE TEMPO, em que “fonte de calor/frio” pode ser

substituído por *fogo / forno / geladeira / freezer / microondas*, e “unidade de tempo” por *minutos / horas / dias / meses*, etc.

- **deixe de molho + PRP + SOLUÇÃO LÍQUIDA**, em que “solução líquida” pode ser substituído por *água (fria) / marinada*, etc.
- **polvilhe com (+ ART) + INGREDIENTE SECO**, em que “ingrediente seco” pode ser substituído por *sal (e pimenta-do-reino) / açúcar (e canela) / QUEIJO ralado / parmesão*, etc.
- **regue com (+ART) + INGR./PREP. LÍQ.** em que “ingrediente/preparação líquidos” pode ser substituído por *azeite, suco (de FRUTA (CÍTRICA)), molho (de tomate / branco / da salada)*, etc.

A grande maioria das formas recuperadas com o padrão “v + adv + adj” começa com um verbo de suporte – *estar(em) / esteja(m) / estiver(em), ficar(em) / fique(m)*. São, na verdade, parte de uma estrutura maior (*quando estiverem bem corados, assim que ficarem macios, até que estejam bem cozidos*, etc.), que deveria ser recuperada por um padrão iniciado por preposição ou pronome, **prp/pron (+ que) v + (+ adv) + adj**, não prevista pela lista inicial.

Por fim, ao final da análise percebi um erro na coleta dos dados – o padrão “v + art + n” havia sido escrito “v + art + NUM” nas expressões de busca, o que provoca uma enorme diferença no número de *types* e *tokens* recuperados. O padrão errado ocorre 185 vezes no corpus, o que rende 9 *types* e 36 *tokens* (nenhum é UTE). A busca utilizando o padrão correto extrai 42.907 ocorrências do corpus, totalizando 2.912 *types* e 29.702 *tokens*. Felizmente, uma análise breve dos dados mostra que há muitas UTEs entre esses *types*, de forma que decidi não repetir todo o procedimento de pesquisa das amostras aleatórias, recalculando os dados das tabelas, refazer os gráficos, etc. após perceber o erro, ainda que as 1.000 ocorrências randômicas analisadas correspondam a apenas 13% do total das formas recuperadas (e não 20%, que foi o mínimo analisado nas outras categorias).

D. Padrões iniciados por artigo

De modo geral, as formas recuperadas iniciadas por artigo não renderam resultados muito interessantes – juntamente com as iniciadas por preposição, foram as

que obtiveram o menor índice de precisão, em parte devido às altas frequências das formas no corpus.

O padrão “art + n + adj” (como em *a parte branca*) recuperou 590 *types* com mais de 2 ocorrências, totalizando 6.654 *tokens*. Mas um exame da lista gerada ordenada alfabeticamente mostra que, em sua maioria, as formas referem-se a ingredientes muito usados (*a cebola picada, a manteiga amolecida, o queijo ralado, etc.*). O que se aproveita, de fato, é pouco: UTEs iniciadas por palavras específicas, como “a parte...” e “o lado...”, além de: art + *INGREDIENTE(S) restante(s)*, art + *INGREDIENTE(S) reservado(s)* e o *máximo / mínimo possível*. As demais, em sua maioria, são pedaços de estruturas maiores, como “(até obter) um creme liso/ claro / homogêneo / brilhante / fofo”.

No caso do padrão “art + n + prp + n”, com 1.215 *types* e 7.727 *tokens*, a proporção de formas recuperadas que não são UTEs parece ser ainda maior. É possível que o uso de uma expressão regular que elimine ocorrências da preposição sem o artigo definido (= “de”) reduza bem o ruído. Enfim, ainda que os padrões iniciados por artigo tenham recuperado UTEs interessantes, como “o restante dos ingredientes”, “a ponta dos dedos”, “o lado da gordura” (vide Anexo V para mais exemplos), acredito que, diferentemente do caso das preposições – em que se pode melhorar a precisão com o uso de *stoplists* – a pesquisa usando o concordanciador para identificar essas UTEs seria uma metodologia mais rápida e eficiente do que o uso de padrões morfossintáticos.

O padrão iniciado por artigo “art + adj + prp + n” foi acrescentado à lista por engano e só recuperou UTEs válidas porque substantivos de alta ocorrência no *corpus* (como *lateral/laterais (da fôrma)*, *fundo (da fôrma)*, *líquido (do cozimento)*, *amarrado (de ervas)*) foram etiquetados como adjetivo.

E. Padrões iniciados por adjetivo

De modo geral, a extração dos padrões iniciados por adjetivo apresentou bastante ruído – ao lado das iniciadas por preposição, foi a categoria que teve o menor índice de precisão. O fator que mais afetou negativamente a extração destes padrões foi a etiquetagem errônea, já mencionada outras vezes, de vários adjetivos de grande ocorrência no corpus. Por outro lado, o fato de outras categorias terem sido etiquetadas

erroneamente como adjetivo chamaram a atenção para estruturas ainda não previstas em outras categorias, como “**prp + adv + v**” (*para não empelotar, para não escurecer*).

Assim como ocorreu com alguns padrões iniciados por substantivos, a extração dos padrões iniciados por adjetivos se beneficiaria do uso de uma *stoplist* para barrar a extração de *types* contendo os adjetivos inespecíficos mais frequentes (como *grande(s)*, *médi(o/a)(s)*, *pequeno(o/a)(s)*, *picad(o/a)(s)*, *cozid(o/a)(s)*, *fin(o/a)(s)*), com exceção de “**adj + adv + adj**” (*picado bem fino*, p. ex.).

Os dois padrões que não obtiveram nenhum *type* com oc. ≥ 3 entre os recuperados foram:

- **adj + prp + adv + n** (proveniente de *frito em bastante óleo*): foi feita uma busca no *corpus* com a palavra “bastante” precedida de preposição e foram recuperadas 137 ocorrências: 60 de “com bastante” e 77 de “em bastante”. No entanto, em nenhuma delas ocorre um adjetivo antes da preposição, apenas verbos (*frite, cozinhe, polvilhe*) e substantivos (*panela, frigideira, bacia*). Isso sugere o acréscimo da estrutura “**v + ... + prp + ... + adv + n**” (*frite em bastante óleo, cozinhe o macarrão em bastante água, polvilhe com um pouco de canela*) aos padrões verbais.
- **adj + adv + adv**: esse padrão entrou na lista por engano, pois estava errado na tabela que originou as expressões (“picado bem fininho” = **adj + adv + ADJ**). Felizmente o padrão correto também faz parte das expressões pesquisadas.

Além desses, o padrão “**adj + prp + adv**” (proveniente de *dourada por cima, rosada por dentro*), com 365 oc. no *corpus*, 28 *types* e 246 tokens também não recuperou nenhuma UTE. Pesquisou-se no *corpus* “por dentro” e “por fora”: há 46 oc. de “por dentro” e 35 oc. de “por fora” – dessas, 22 são oc. de “por dentro e por fora” (há apenas 1 oc. de “por fora e por dentro”). A categoria que mais co-ocorre do lado esquerdo da expressão é, de fato, o adjetivo, mas há muita variação. À esquerda de “por fora” há, por exemplo: *dourado, dourada, dourados, douradas* (5 oc. no total); à esquerda de “por dentro”: *cozido, cozida, cozidos, cozidas* (6 oc. no total) e *rosado, rosados, rosada, rosadas* (5 oc. no total). À esquerda de “por dentro e por fora” há vários verbos e sintagmas verbais do campo semântico TEMPERAR (*salpique, tempere, esfregue*, etc.). Não há outras UTEs entre os *types* recuperados, de forma que o padrão poderia ser reescrito como **adj + “por” + “fora/dentro” (+ “e por fora/dentro”**.

Os padrões que renderam mais UTEs foram “**adj + adv**” (picadas grosseiramente, fatiado finamente, ralado grosso) e “**adj + conj + adj**”, que rendeu muitos binômios (*lisa e elástica, dourada e crocante, quente ou frio*) – alguns deles já haviam sido identificados em estruturas maiores, mas pelo fato de estarem separados dessas estruturas aqui faz com que a extração automática seja mais eficiente.

F. Padrões iniciados por advérbio

Os padrões iniciados pela categoria advérbio ficaram atrás apenas dos iniciados por verbos no que diz respeito à precisão na extração de UTEs do corpus, com 48% de formas válidas nas amostras aleatórias. Foi uma das categorias iniciais que teve *types* com frequências mais altas (ou seja, um alto nível de repetição) – 22% dos *types* recuperados (= 838) são responsáveis por 74% (= 10.024) de todas as ocorrências dos padrões no corpus. Além disso, é a única cujos padrões, sem exceção, recuperaram pelo menos um *type* no *corpus* com 3 ou mais ocorrências. E todos recuperaram pelo menos uma UTE – ainda que o padrão “adv + v + prp + v” (proveniente de *sem parar de mexer*) não tenha recuperado nenhuma UTE, devido a um erro na etiquetagem da UTE original, já que “sem” é preposição e não advérbio.

Quanto à eficiência dos padrões individuais, observou-se que no padrão “adv + adj”, por exemplo, seria vantajoso eliminar as formas iniciadas por “muito” e “pouco”, mas não sem antes fazer um estudo da prosódia semântica de “muito”, que coleciona contextos negativos na Culinária nos casos em que é precedido de “não”, (padrão “adv + adv + adj”): *não muito cheia(o/a)(s), dur(o/a)(s), escur(o/a)(s), espess(o/a)(s), fin(o/a)(s), forte(s), grande(s), gross(o/a)(s), líquido(o/a)(s), maci(o/a)(s), mole(s), pegajos(o/a)(s)*, etc.

No caso do padrão “adv + conj + v”, somente as formas iniciadas por “assim que” são UTEs válidas – o padrão poderia ser reescrito para recuperar formas como *assim que a água ferver, assim que o molho começar a engrossar, assim que levantar fervura*, etc. Já em “adv + v”, seria vantajoso eliminar as iniciadas por “mais”, “depois” e “também”.

Alguns padrões recuperaram poucas UTEs, mas muito recorrentes: “adv + conj + adv” (*mais ou menos*, 60 oc.); “adv + pron” (*enquanto isso*, 301 oc.), “adv + prp + num + n” (*cerca de X minuto(s)* (877 oc.) / *hora(s)* (67 oc.) / *cm* (152 oc.) / *kg* (61 oc.), etc.). O padrão “adv + n”, por sua vez, recuperou 90 *types* com oc. ≥ 3, mas pouquíssimas

UTES – e, observando o resultado *en passant*, parece que não é possível usar regras ou *stoplist* para melhorá-la, pois o resultado é heterogêneo. Seria o caso, talvez, de eliminar o padrão da lista.

4.2.5.5 Resumo das estratégias para melhorar a extração

Em vista do que foi observado nos resultados obtidos na extração semi-automática de candidatas a UTE no *corpus* de culinária usando padrões morfossintáticos, pode-se concluir que é possível tomar várias medidas de precaução antes da extração e durante a mesma para evitar ruído nos dados recuperados, que dificulta a análise da precisão do método:

- Os números do *corpus* devem ser transformados em uma única variável para unificar os padrões que o contêm.
- A busca por padrões deve ignorar maiúsculas e minúsculas, ou pode-se trabalhar com uma versão do *corpus* em que as letras sejam todas transformadas em minúsculas (ainda que se percam, nesse caso, informações importantes como nomes / marcas de produtos, lugares, etc). Talvez o ideal fosse usar um etiquetador que tivesse um bom repertório de entidades nomeadas, ou seja, uma lista de lugares, nomes, marcas, etc. que, mesmo tendo mais de um item lexical em sua composição, são consideradas com uma unidade (p. ex., “Universidade de São Paulo” = n). É claro que isso afetará a maneira como as expressões de busca são construídas.
- O uso de etiquetas pragmáticas, como as que foram empregadas em meu mestrado (Teixeira, 2004, cap. 1), na preparação do *corpus* quando de sua compilação teria sido de grande valia no caso específico da culinária, especialmente para melhorar a coleta de padrões que contém a categoria “substantivo”. Seria interessante, por exemplo, etiquetar todas as informações da receita que não são estritamente parte da lista de ingredientes e/ou do modo de fazer – inclusive título, informações nutricionais, tipo de cozinha, tempo de preparo, grau de dificuldade, etc. – para que possam ser eliminadas (ou pesquisadas separadamente, caso seja de interesse) nas buscas por padrões.
- As buscas por padrões devem ser capazes de ignorar formas que contêm não apenas sinais de pontuação em seu interior (o que é relativamente fácil de fazer

quando se usa uma regra, por exemplo, que permita apenas os parênteses no meio de um padrão – “colher (chá)” – mas não ponto final, ponto-e-vírgula, etc.), mas também quebras de linha, de parágrafo, de seção, etc. Isso pode ser conseguido por meio de etiquetagem, por exemplo.

- Pelo que observei nos dados coletados, o número mínimo de ocorrências para padrões com 4 ou mais constituintes pode ser diminuído para 2.
- Pode-se usar *stoplists*, quando necessário, mas elas devem ser específicas para padrões ou grupos de padrões, ou para posições distintas (começo / fim), e não como regra geral em uma determinada categoria ou das categorias como um todo.
- A escolha de um etiquetador que dê conta, o máximo possível, das particularidades da área especializada em estudo, e a correção posterior das etiquetas mais recorrentes do *corpus* é imprescindível para o sucesso da extração.
- O estudo minucioso da lista de padrões que será usada na extração e a testagem prévia desses padrões em uma porção aleatória do *corpus* é essencial para o sucesso da metodologia de coleta. Expressões regulares aptas a prever o máximo possível de variações que uma estrutura pode ter minimizam o número de expressões de busca e de resultados indesejados.

4.3 Considerações finais sobre a extração semi-automática de UTEs

Neste capítulo foram feitos vários estudos, usando diversos procedimentos da Lingüística de Corpus, para observar a natureza e o comportamento das Unidades de Tradução Especializada em receitas culinárias. O objetivo era incorporar esse conhecimento a métodos de extração semi-automatizados, de forma que fossem mais eficientes na recuperação de unidades desse tipo em *corpora* especializados.

No primeiro estudo, foi feito um levantamento da padronização da palavra-chave “flour” num sub*corpus* de inglês de cerca de 250.000 palavras usando o programa WordSmith Tools. Em seguida, foi feito um levantamento, quando possível, da frequência desses padrões na Internet usando o Google – em sua grande maioria, os padrões mostraram-se muito recorrentes em textos da *web*. Depois, os mesmos padrões foram pesquisados em vários dicionários da área, bilíngües e monolíngües,

disponíveis no Brasil. A frequência dos equivalentes sugeridos (quando havia algum) foi verificada na Internet, mais uma vez, usando o Google.

A discrepância entre o número de padrões levantados no *corpus* e o número de verbetes nos dicionários analisados contendo o radical pesquisado confirmou a hipótese de que os dicionários bilíngües e multilíngües disponíveis para o tradutor da área no par de línguas português-inglês não possuem informações suficientes para auxiliá-lo na produção textual na língua de chegada. As obras focam quase que exclusivamente os substantivos, preocupando-se apenas com a operação de decodificação do sentido. Não há rigor metodológico na elaboração das entradas, não são oferecidos exemplos de uso, e alguns equivalentes sugeridos são pouco ou nada usados em textos reais da área.

No segundo estudo, foram analisados cerca de 1.300 entradas (UTES) em português provenientes do glossário português-inglês que compilei em minha memória de tradução de Culinária quanto a: i) categoria gramatical de cada palavra; ii) função sintática da entrada como um todo; iii) número de elementos; iv) presença ou não de palavras-chave; v) frequência no *corpus* de português total, com cerca de 1,5 milhão de palavras.

Com relação à categoria gramatical, observou-se que, apesar de a grande maioria das entradas constituir-se de ou começar com substantivos (68%) e verbos (18,9%) – o que era de se esperar –, mais de 10% das entradas começa ou constitui-se de categorias em geral desprezadas na compilação de dicionários técnicos: adjetivos (5,5%), preposições (5,5%), advérbios (1,2%) e outras categorias, como conjunção, pronome relativo e determinante (menos de 1%) – essas porcentagens tornam-se ainda mais expressivas quando se considera o número de substantivos ocorrendo sozinhos nesta lista (325 – 25%). Ainda com relação ao número de elementos das UTES, constatou-se que as formações sintagmáticas (827 oc., a grande maioria com 2 a 5 elementos) são duas vezes mais frequentes na lista de entradas do que categorias gramaticais ocorrendo sozinhas (467 oc.) – o fato de eu ter dedicado mais tempo à inclusão das mesmas no glossário da memória de tradução ao traduzir pode ser uma evidência de que são mais importantes para a produção textual na língua de chegada e/ou de que são mais difíceis de serem localizadas nos materiais disponíveis e/ou na Internet.

No que respeita à presença de palavras-chave nas UTEs, apenas 113 (8,7%) não passaram no teste; 107 (95%) delas contêm apenas um elemento, muitas das quais são provenientes de um glossário de ervas e especiarias preparado previamente para uma disciplina de Terminologia. A presença de “enquanto isso” nesta lista chamou minha atenção para o perigo de se usar somente esse critério na identificação de candidatos a UTE no *corpus* – embora as palavras “enquanto” e “isso” isoladas não sejam chave, a combinação é. Quanto à frequência, 137 (10,6%) UTEs tiveram ocorrência igual a zero no *corpus* de português – 55 (40%) delas são monovocabulares e estão também na lista de UTEs sem palavra-chave. Chama a atenção também, nesta lista, a presença de termos raros, mas não em desuso na culinária (como “feno-grego” e “macis”) – o aumento do *corpus* seria suficiente para garantir sua presença nos textos da área?

Tendo concluído essa análise inicial das UTEs de um ponto de vista mais semântico-lexical, passei à última etapa do estudo prático, que aborda essas unidades pelo viés morfossintático. O objetivo era divisar metodologias semi-automáticas para a extração de padrões semelhantes em *corpora* técnicos. Para tanto, tomei os padrões morfossintáticos levantados no segundo estudo (cerca de 130) e, usando métodos de extração semi-automáticos, foram recuperadas as formas com padrão idêntico contidas na versão previamente etiquetada do *corpus* total de português. Uma amostra aleatória do resultado foi analisada para determinar a precisão do método.

Observou-se, entre outras coisas, que: i) a precisão na etiquetagem do *corpus* é um dos fatores mais importantes no sucesso das metodologias de extração por padrão morfossintático – é necessário, portanto, certificar-se de que as etiquetas são adequadas aos textos da área especializada antes de empreender as buscas; ii) o uso de lematização e a transformação de números em variáveis teriam tornado a coleta mais eficiente, evitando a omissão de itens importantes; iii) o uso de *stoplists* específicas para determinadas categorias, por outro lado, diminuiria a quantidade de ruído nos resultados obtidos.

O exame geral dos dados obtidos nessas pesquisas permite afirmar que:

1. devido à natureza formulaica das receitas, muitas UTEs são maiores do que os dicionários técnicos geralmente consideram como “entradas” – para que separar o padrão “abra a massa PREPOSIÇÃO + ARTIGO INDEFINIDO superfície (lisa e) enfarinhada / polvilhada com farinha” em pequenas unidades (p. ex.: “abrir”,

“massa”, “superfície”, “lisa”, “polvilhar” e “farinha”) se essa UTE ocorre, com pequenas variações, quase 100 vezes no corpus? Ou seja, ela é uma Unidade de Tradução para a qual o tradutor terá de encontrar uma forma equivalente;

2. a padronização em textos técnicos vai muito além dos substantivos e dos verbos: os advérbios e as preposições, por exemplo, são muito freqüentes nos padrões da área, desempenhando um papel importante na convencionalidade e, portanto, na produção textual na língua de chegada. Foram identificados também vários padrões que não contêm “termos”, no sentido estrito (unidades conceptuais), mas que são freqüentes e importantes na padronização textual;
3. nos casos de padrões com muitas unidades (acima de 4 ou 5) parece haver uma associação muito mais forte entre itens lexicais do que entre itens lexicais e uma determinada estrutura morfossintática rígida; ou seja, é possível que a consideração das associações entre itens lexicais seja mais importante do que a consideração de estruturas morfossintáticas na busca por padrões em textos técnicos;
4. ambos os métodos testados – análise “manual” de todas as concordâncias de uma palavra-chave e extração semi-automática por meio de padrões morfossintáticos – mostraram-se falhos. O primeiro pela morosidade, o segundo pela quantidade de ruído presente nos resultados recuperados no corpus. O ideal seria desenvolver um método que associasse essas duas abordagens, como já vem sendo feito por alguns pesquisadores da área de PLN para a língua geral, mas respeitando as especificidades discursivas das áreas técnicas. Enquanto isso não se concretizar, o método manual, apesar da morosidade, me parece o mais eficiente, já que o segundo demanda um conhecimento de computação, cuja aquisição pode ser ainda mais morosa, para um lingüista, do que analisar uma a uma as linhas de concordância para cada palavra-chave dos *corpora* de cada uma das línguas.

Em suma, essas conclusões apontam para o fato de que a cooperação entre tradutores, terminólogos e lingüistas de corpus, de um lado, e desses com os pesquisadores em PLN, de outro, é essencial para avançar os estudos nessa área.

Para tentar disponibilizar para o tradutor os resultados obtidos nessas pesquisas, bem como os conhecimentos teóricos que os embasam, proponho um

protótipo de verbete para um Dicionário *Online* de Culinária, apresentado no capítulo a seguir.

This document was created with Win2PDF available at <http://www.daneprairie.com>.
The unregistered version of Win2PDF is for evaluation or non-commercial use only.

Capítulo V

Proposta de um Dicionário *Online* de Culinária voltado para o tradutor

Neste capítulo, apresentamos a proposta para um dicionário *online* de Culinária voltado para o tradutor que contemple, em vários pontos de sua micro- e macroestrutura, as necessidades de produção textual deste profissional. Inicialmente, discutimos os elementos diferenciais da interface *online* que abrigará o dicionário e as implicações desse tipo de interface para a macroestrutura da obra e para o usuário. Em seguida, propomos o modelo de verbete voltado para o tradutor, caracterizando o tipo de entrada e de paradigma informacional, pragmático e de forma equivalente que comporão a microestrutura. Por fim, exemplificamos com dois verbetes contendo uma UTE em cada direção tradutória.

5.1 Por que um Dicionário de Culinária *online*

Dicionários eletrônicos têm muitas vantagens para o consulente, se comparados aos dicionários impressos – a nosso ver, as duas únicas qualidades que se poderiam destacar nestes últimos são a possibilidade de serem consultados sem a necessidade do computador e a simultaneidade visual de vários verbetes, possivelmente relacionados, numa mesma página. Os dicionários eletrônicos, por seu turno, destacam-se dos impressos em vários aspectos, dentre eles:

- **a consulta é mais rápida:** em geral, basta digitar a palavra procurada numa caixa de diálogo para ter acesso ao verbete;
- **a busca é mais refinada:** é possível fazer uma pesquisa reversa, por exemplo, apenas no campo *etimologia*, ou *definição*, de alguns dicionários eletrônicos, sem mencionar o fato de que é muito mais fácil e rápido navegar de um verbete a outro (em muitos dicionários, o uso de *hyperlinks* faz com que um simples clique sobre a palavra desejada conduza o usuário diretamente para o verbete);
- **não há limitações de espaço:** como o dicionário eletrônico não tem as limitações de tamanho impostas pela versão em papel, pode conter muito mais informações (p. ex.: a conjugação de verbos em todos os tempos e modos verbais nos dicionários de língua portuguesa), além de conteúdos multimídia (sons, vídeos, esquemas animados, etc), recursos incompatíveis com as obras impressas;
- **o aspecto visual é mais bem trabalhado:** o custo de recursos gráficos, como texto em cores diferenciadas para cada tipo de informação do verbete, imagens, quadros de destaque com fundo colorido, etc. é muito alto nas obras impressas, mas em obras eletrônicas não – isso faz com que o aspecto visual de dicionários eletrônicos seja mais bem definido, facilitando a busca e tornando-a mais rápida e eficiente.

Os dicionários *online* apresentam todas as vantagens do dicionário eletrônico e ainda oferecem outros diferenciais: i) permitem a atualização e o aperfeiçoamento constante do conteúdo disponibilizado; ii) permitem a integração de vários outros recursos também disponíveis *online*, oferecendo a possibilidade de concentrar, num único ambiente, várias possibilidades para uma consulta completa e variada sobre uma determinada palavra, expressão ou tema; iii) permitem a interatividade com o usuário,

possibilitando conhecer suas necessidades para atendê-las de forma mais rápida e eficiente. A única desvantagem do dicionário eletrônico *online* para o usuário é o fato de a consulta estar condicionada ao uso do computador e à disponibilidade de uma conexão com a Internet.

Entretanto, conforme demonstrou Fromm (2008), a Internet é uma das ferramentas de pesquisa mais utilizadas pelos tradutores atualmente. Esses profissionais trabalham com várias fontes de consulta simultaneamente – não necessariamente por não haver uma obra que concentre todas as informações de que necessitam para produzir uma boa tradução, mas porque a variedade de fontes é percebida como um dos fatores que ajuda a garantir a acurácia e idoneidade da informação pesquisada.

Pensando no perfil do tradutor, nas necessidades apontadas em várias pesquisas da área, e levando em conta nossa própria experiência no ensino e na prática de tradução, planejamos o Dicionário Online de Culinária para Tradutores (doravante DOC-Trad), que passamos a descrever no item abaixo.

5.1.1 Macroestrutura: a interface *online* do Dicionário Online de Culinária

O DOC-Trad será disponibilizado *online* e o *website*²³⁰ que o abrigará conterá os seguintes *links* no menu lateral à esquerda da página (vide Fig. 73):

- **Página inicial:** conterá *links* que conduzem à descrição da proposta do *site*, um histórico das atualizações, equipe de trabalho, parceiros e colaboradores, um *link* para o usuário se inscrever e receber notícias sobre as últimas atualizações (que poderão ser enviadas, por exemplo, via *blog*), um fórum de dúvidas (que poderá servir de guia para futuras atualizações do dicionário, com base nas necessidades e dúvidas dos usuários) e o endereço de contato para envio de críticas e sugestões (estes dois últimos também constam da lista de *links* do menu lateral, como pode ser visto na figura 73);
- **Tradução Culinária:** conterá informações sobre aspectos da tradução no gênero Culinária e seus tipos textuais mais típicos, em especial as receitas (adaptação das informações que constam do Capítulo II);

²³⁰ www.traducaoculinaria.com.br/DOC-Trad

Home	<p>→ this site in English</p> <p>DOC-Trad Dicionário Online de Culinária para Tradutores português-inglês / inglês-português</p> <p>A proposta deste site</p> <p>Histórico das atualizações</p> <p>Equipe de trabalho</p> <p>Parceiros e colaboradores</p> <p>Receba notícias sobre as últimas atualizações</p> <p>Participe do fórum de dúvidas</p> <p>Envie suas sugestões, críticas e elogios</p>
Tradução Culinária	
Como usar o DOC-Trad	
Consultar DOC-Trad (por palavra)	
Consultar DOC-Trad (por palavra-chave)	
Consultar DOC-Trad (por área)	
Consultar CorTrad (corpus paralelo)	
Consultar CorTec (corpus comparável)	
Pesquisar no Google:	
usar o KWIC Google	
Consultas externas	<p>última atualização: XX/XX/XXX</p> <p>contador de acessos: XXX</p>
Fórum de dúvidas	
Contato	

Figura 73: página inicial do Dicionário *Online* de Culinária para Tradutores

- **Como usar o DOC-Trad:** conterá explicações sobre a estrutura do verbete e do dicionário, dando especial ênfase aos recursos diferenciais da obra – lista de palavras-chave, linhas de concordância, comentários para o tradutor, frequências e árvore hiperbólica (vide item *Consultar dicionário (por área)* a seguir); também dará instruções básicas de como e em que situações consultar as ferramentas do CorTec, CorTrad e KWIC Google (que também serão *links* do menu lateral – vide a seguir);
- **Consultar DOC-Trad (por palavra):** listará todas as palavras contidas em quaisquer campos do dicionário, em ordem alfabética, que, ao serem clicadas, exibem uma lista de todos os verbetes em que a palavra clicada aparece pelo menos uma vez em qualquer parte do verbete (entrada, exemplo, comentário, etc.). Trata-se, na verdade, de uma espécie de “wordlist” do dicionário, que funcionará como um índice remissivo por palavra, e não por assunto; também permite a busca por meio da digitação da palavra de busca numa caixa de diálogo (com a possibilidade de uso de caractere coringa);

- **Consultar DOC-Trad (por palavra-chave):** listará as palavras-chave do corpus por ordem decrescente de chavicidade (da mais chave para a menos chave), podendo ser consultada também pela ordem alfabética – como ilustra a figura 74, ao clicar sobre uma palavra-chave, o usuário tem acesso a uma lista com todas as entradas que contém aquela palavra chave ou uma categoria semântica que a englobe (p. ex.: ao clicar na palavra-chave “azeite”, além de “azeite de oliva virgem”, “azeite-de-dedê”, etc. os itens “fio de ÓLEO”, “regar com ÓLEO”, “aqueça um fio/pouco de ÓLEO numa PANELA”, “ALIMENTO conservado em ÓLEO”, etc. também serão listados);

Home		
Tradução Culinária	1. SOPA	→ in English
Como usar o DOC-Trad	2. SAL	→ ver em ordem alfabética
Consultar DOC-Trad (por palavra)	3. CHÁ	
Consultar DOC-Trad (por palavra-chave)	4. COLHERES	
Consultar DOC-Trad (por área)	5. PREPARO	
Consultar CorTrad (corpus paralelo)	6. XÍCARA	
Consultar CorTec (corpus comparável)	7. INGREDIENTES	
Pesquisar no Google:	8. COLHER	
usar o KWIC Google	9. COLOQUE	
Consultas externas	10. MANTEIGA	
Fórum de dúvidas	11. MINUTOS	
Contato	12. MISTURE	
	13. RETIRE	
	14. DEIXE	
	15. PANELA	
	16. FOGO	
	17. FORNO	
	18. FARINHA	
	19. AZEITE	
	• azeite (de oliva)	
	• azeite(-)de(-)dedê	
	• azeite (de oliva) virgem	
	• azeite (de oliva) extra(-)virgem	
	• fio de ÓLEO	
	• regue com um fio de ÓLEO	
	• aqueça um fio de ÓLEO numa PANELA	
	• etc.	
	20. ÁGUA	
	21. etc...	

Figura 74: consulta do DOC-Trad por palavra-chave

- **Consultar DOC-Trad (por área):** futuramente, abrigará uma representação visual dos campos semânticos da culinária e, especialmente, das receitas, do ponto de vista textual: as palavras que ocorrem tipicamente nos títulos, nas listas de ingredientes, no modo de preparo, nos comentários, etc. Como esse tipo de informação não pôde ser levantado no corpus como o que foi construído para a presente pesquisa (por não termos conseguido identificar essas partes do texto automaticamente por meio da colocação de etiquetas), fizemos apenas um

paralelos de culinária compilados para esta pesquisa para os quais obtivemos autorização de uso, mas que não tivemos tempo hábil de alinhar para disponibilização;

- **Consultar CorTec:** direcionará o usuário para a página de pesquisa do CorTec²³⁵, corpus técnico comparável desenvolvido no âmbito do Projeto COMET que permite pesquisar, entre outros, o Corpus Comparável de Culinária Português-Inglês compilado para a presente pesquisa; lá, o usuário poderá examinar outros contextos de uso para os itens lexicais pesquisados, além de *clusters* (n-gramas) e da lista de palavras dos *corpora* em ambas as línguas;
- **Pesquisar no Google:** direcionará o usuário para a página de buscas no Google;
- **Usar KWIC Google:** direcionará o usuário para a página do KWIC Google, ferramenta criada pelo Prof. Tony Berber Sardinha, no âmbito do LAEL da PUC-SP, que permite colar o resultado de uma pesquisa feita no Google e gerar linhas de concordância com esse resultado a partir de uma palavra de busca²³⁶;
- **Outras fontes de pesquisa:** oferecerá vários *links*, em português e inglês, para *sites* disponíveis na web relacionados à Culinária (p. ex.: dicionários e glossários de culinária, *sites* de receitas de fontes confiáveis, *sites* em inglês sobre culinária brasileira, etc.) e à pesquisa lingüística na Internet (concordanciadores que usam a *web* como corpus, motores de busca, dicionários monolíngües e bilíngües, etc.). A lista será constantemente atualizada (vide Fig. 76).

²³⁵ http://www.fflch.usp.br/dlm/comet/consulta_cortec.html [acesso em 05/09/2008]

²³⁶ para mais detalhes sobre o funcionamento do KWIC Google, vide item 2.4.2.3 (p. 142).

Home	<p>Outras fontes de pesquisa:</p> <p>Localização de informações na Internet:</p> <ul style="list-style-type: none"> • Google – aprenda a tirar melhor proveito do Google para a tradução lendo o tutorial do <i>site</i> sobre suas ferramentas avançadas, ou no site da UC Berkley (http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/Handouts.html); pesquisa apenas em livros publicados (http://books.google.com/) [principalmente inglês]; pesquisa de imagens (http://images.google.com/) • WiseGeek (www.wisegeek.com/) – site com definições retiradas da web (muitas delas revisadas), em diversas áreas, inclusive culinária [só inglês] • Clusty (http://clusty.com/) – faz uma busca rápida das associações mais comuns à palavra / expressão de busca na web, dividindo as respostas em "categorias", que variam conforme a busca (por ex: para <i>banana</i> as "categorias" apresentadas incluem: recipes, tree, fruit, leaf, Banana Republic, etc.) • WebAsCorpus (http://webascorpus.org/searchwac.html) – website que usa a web como corpus – os resultados podem ser visualizados como linhas de concordância, com a palavra / expressão de busca em destaque • (etc.) <p>Dicionários:</p> <ul style="list-style-type: none"> • OneLook (http://www.onelook.com/) – vários dicionários monolíngües e multilíngües, tradutores automáticos, motores de busca, etc. que podem ser consultados simultaneamente [principalmente inglês] • WordReference (http://www.wordreference.com/enpt/) – ainda em construção, é um dicionário bidirecional português-inglês (tem versões em outros pares de língua) com exemplos de uso • (etc.) <p>Sites com informações Culinárias:</p> <ul style="list-style-type: none"> • SIC – Serviço de Informação da Carne (http://www.sic.org.br/ilustracao.asp) – site bilíngüe com informações sobre cortes de carne bovina, com fotos, descrição dos usos culinários e receitas • Oliver's Artisan Bread (http://oliversbreads.com/breadtypes.html) – site de uma padaria americana com fotos de vários formatos de pão tradicionais [em inglês] • Cook's Thesaurus – dicionário de ingredientes dividido por campo semântico mas que permite a pesquisa por termo; com fotos, descrição, variantes denominativos, sugestões de substituição, etc. [inglês] • (etc.)
Tradução Culinária	
Como usar o DOC-Trad	
Consultar DOC-Trad (por palavra)	
Consultar DOC-Trad (por palavra-chave)	
Consultar DOC-Trad (por área)	
Consultar CorTrad (corpus paralelo)	
Consultar CorTec (corpus comparável)	
Pesquisar no Google:	
Usar o KWIC Google	
Consultas externas	
Fórum de dúvidas	

Figura 76: links para outros sites relacionados à Culinária ou à pesquisa na Web

Essa estruturação da interface possibilitará o acesso total, não somente às entradas, mas aos contextos de uso (traduzidos ou não), por meio dos vários tipos de consulta oferecidos – no dicionário de papel, muitas vezes, a informação não pode ser localizada com facilidade, ainda que conste de outros verbetes relacionados ao que o consulente está pesquisando.

O maior diferencial da macroestrutura do DOC-Trad, no entanto, é que esta não se organiza a partir da árvore de domínio da área especializada, como é feito nos dicionários terminológicos, mas a partir da lista de palavras-chave do corpus de estudo. Isto é, o ponto de partida para o trabalho de recolha de UTEs é a lista de palavras-chave de cada língua. A partir das palavras com maior chavicidade no corpus são geradas linhas de concordância e, a partir da recolha manual (observando as linhas de concordância uma a uma) ou semi-automatizada (conforme exemplificado no Capítulo IV), são identificados os padrões que comporão a lista de entradas. O mesmo é feito com a(s) palavra(s)-chave da língua de chegada que poderia(m) ser considerada(s) equivalente(s) *prima facie* (identificado por meio da consulta feita nos dicionários disponíveis, no corpus paralelo, no corpus comparável ou em outras fontes de informação). Somente após o levantamento dos padrões na língua de chegada é que

são estabelecidas as possíveis equivalências entre as UTEs. Quando não for possível identificar no corpus da língua de chegada um equivalente de uso consagrado, pode-se lançar mão da consulta junto a especialistas, para tentar oferecer uma sugestão de tradução ou de informação que possa ser usada numa nota explicativa ou de rodapé (p. ex., como traduzir “green walnut preserve” ou “chicken is ready when the thermometer reads 185 F”).

5.1.2 A ficha para registro dos dados

Os dados que comporão o DOC-Trad serão armazenados num banco de dados (inicialmente construído em Access) cuja ficha conterá os seguintes campos:

Campo	Descrição
UTE_LP	Obrigatório – unidade de tradução na língua de partida.
def_LP	Inativo – poderá ser usado para abrigar uma definição na LP, no caso de termos, etapa não contemplada neste trabalho, mas prevista para uso futuro.
freq_LP	Obrigatório – frequência da UTE_LP no corpus da LP; tentaremos usar um mecanismo que extraia esse dado automaticamente do corpus; assim, conforme o corpus for modificado, não será preciso atualizar os dados manualmente no banco.
img_main	Opcional – imagem ou vídeo referente à entrada ou algum aspecto que ela englobe ou de que participe; algumas fraseologias não possuirão imagem ou vídeo (como “enquanto isso”), caso em que o campo ficará em branco.
img_xtra	Opcional – lista com outras imagens disponíveis em nossos arquivos (idealmente, abrirão todas numa nova página, com a possibilidade de seleção e zoom, etc).
link_GI	Fixo – direciona o usuário para a página do Google Image Search, onde ele pode fazer outras buscas por imagens
link_vds	Fixo – lista com <i>sites</i> que disponibilizam vídeos gratuitos ilustrando técnicas culinárias.
UTE_LCh_1	Obrigatório – equivalente(s) na língua de chegada. Quando não houver um equivalente consagrado na língua de chegada, será colocado “vide comentários para o tradutor”; quando o equivalente for uma sugestão de tradução (termo não consagrado, pouco usado, etc.), será marcado com algum recurso gráfico (com asterisco, de outra cor, em itálico) – em ambos os casos, o campo coment (comentários para o tradutor) torna-se de preenchimento obrigatório.

freq_LCh1	Obrigatório – frequência da UTE_LCh no corpus da língua de chegada; também será, idealmente, extraída automaticamente do corpus; quando não houver equivalente, o valor “-” será atribuído automaticamente ao campo.
UTE_LCh2...n	Opcionais – outros possíveis equivalentes, na língua de chegada, para a entrada a UTE que é entrada do verbete. Para cada sinônimo acrescentado, o programa oferece a opção de abrir uma nova ficha na direção em questão.
freq_LCh2...n	Condicional (idem freq_LCh1) – frequência da UTE_LCh2 no corpus; só será preenchido se houver mais de um equivalente na língua de chegada.
coment	Opcional – associado à UTE_LP e escrito na língua de chegada, conterá comentários voltados para o tradutor, enfatizando a produção textual em LCh.
ex_1	Obrigatório – primeiro dos 3 exemplos que consistem em um contexto associativo produzido naturalmente na LP alinhado com sua respectiva tradução na LCh; idealmente, poderão ser extraídos dos corpora paralelos ou, em último caso, do corpus comparável da língua de partida e, então, traduzidos.
ex_2	Obrigatório – exemplo 2 (idem acima).
ex_3	Obrigatório – exemplo 3 (idem acima).
CorTrad_10	Obrigatório – até 10 contextos paralelos fixos (LP e respectiva tradução em LCh) que serão extraídos do CorTrad. O tradutor terá acesso opcional a esses contextos, que serão mostrados no mesmo local (expandindo, sem abrir nova página), por meio de um link presente na página do verbete. Caso não haja nenhum contexto paralelo no CorTrad para o item lexical buscado, será exibida a mensagem: "não há contextos contendo o item pesquisado no Corpus Paralelo de Culinária do CorTrad".
CorTec_10	Obrigatório – até 10 contextos monolíngües extraídos do corpus comparável originalmente escrito em LCh (se houver mais de um equivalente em LCh, serão incluídos contextos com esses equivalentes também). Assim como o campo acima, será acessado via link (expandindo, sem abrir outra página).
fconc_1...n	Opcional – formas concorrentes (sinônimos) da UTE na língua de partida.
freq_sin1...n	Condicional – frequência da forma concorrente no corpus (idealmente, será extraída automaticamente, para que se atualize junto com o corpus).
ref_cruz	Opcional – outras entradas relacionadas à UTE_LCh; serão links que, ao serem clicados, darão acesso àquele verbete (todos os itens que constarem deste campo devem possuir verbete próprio no dicionário).
loc_repr	Inativo – não entrará neste trabalho, mas servirá para abrigar o

“endereço” da UTE na representação visual dos campos semânticos da área (árvore hiperbólica).

Tabela 19: campos da ficha de recolha de dados

5.1.3 Microestrutura: o verbete

O termo *microestrutura* refere-se ao modo como os dados contidos no verbete organizam-se, isto é, como os paradigmas (informacional, definicional, pragmático e de equivalência) são dispostos no enunciado terminográfico (Barros, 2004: 156). O verbete do DOC-Trad, como discutido anteriormente, não possui paradigma definicional, por enquanto, embora esteja reservado um espaço para essa informação na ficha de recolha de dados, conforme pode ser visto na Tabela 19. Nos itens que se seguem, apresentamos a entrada e os demais paradigmas do verbete mostrando como estes buscam suprir as carências que identificamos nos dicionários bi- e multilíngues de culinária no que se refere à produção textual.

5.1.3.1 Entrada

A lista de entradas do DOC-Trad consistirá das UTEs identificadas no corpus comparável, isto é, os padrões léxico-gramaticais identificados em textos especializados originalmente produzidos em cada uma das línguas do glossário por meio de metodologias de exploração de corpora desenvolvidas no âmbito da L.C., como as descritas no Capítulo IV. As entradas do DOC-Trad apresentam alguns diferenciais em relação aos dicionários técnicos disponíveis atualmente:

- **tamanho:** no geral, são seqüências mais longas de palavras, p. ex.: “asse até que, enfiando um palito no meio, ele saia limpo”; “vire a massa sobre uma superfície (ligeiramente) polvilhada (com farinha)”;
- **presença de categorias semânticas:** quando for identificada no corpus a ocorrência de um padrão léxico-gramatical que aceite várias palavras de uma mesma categoria semântica ou gramatical, esta será representada na entrada por letras maiúsculas colocadas entre colchetes (p. ex.: “gomo de [FRUTA CÍTRICA]”) e, no comentário para o tradutor, serão listadas as palavras da categoria semântica que mais ocorrem naquela posição (p. ex.: “laranja(s)” (60 oc.), “limão” (40 oc.), “tangerina” (14 oc.));

- **ordem:** como a consulta ao dicionário pode ser feita de várias formas (ordem alfabética de todas as palavras, por palavra-chave, por busca com uso de caractere coringa), a ordem dos itens lexicais na entrada seguirá sempre a ordem natural de ocorrência no corpus, p. ex.: “(red) kidney bean” (e não “bean, (red) kidney”); “bata as claras em neve” (e não “claras em neve, bater”);
- **forma marcada:** as entradas não são necessariamente escritas na forma não-marcada (infinitivo para verbos, masculino singular para substantivos e adjetivos, etc.), mas na forma que ocorrem com maior frequência no corpus naquele padrão específico, p.ex.: “tempere a gosto com sal e pimenta-do-reino” (e não “temperar a gosto...”); “bata as claras em neve” (e não “bater clara em neve”), mas “[INGREDIENTE] picado” (representando “picados”, “picada” e “picadas”).

5.1.3.2 Paradigma informacional

Como a maioria das entradas consiste de formações sintagmáticas e frases mais longas, o DOC-Trad não apresenta informações sobre sua categoria gramatical.

O maior diferencial do paradigma informacional do DOC-Trad está no “**Comentário para o tradutor**”, em que serão dadas informações que visam auxiliar o profissional ou aprendiz especificamente com relação à produção textual na língua de chegada, p. ex.: “nos EUA, em geral, usa-se um termômetro para testar a temperatura interna das carnes e determinar seu ponto de cozimento”; “*seedless* é usado para indicar alimentos alterados geneticamente de modo a não conterem sementes, ao passo que *seeded* é usado para indicar a remoção das sementes de um alimento que as possui naturalmente”; “o tamanho mais difundido da lata de feijão nos EUA é 14 oz (410g)”, etc.

Para os verbetes em que isso for possível, será apresentada uma imagem ilustrativa ou vídeo no canto superior direito da página, com três *links* logo abaixo da imagem: um apontando para outras imagens/vídeos de nosso arquivo próprio, um apontando para a página inicial do *Google Image Search*, em que o usuário poderá pesquisar imagens disponíveis na *web*, e outro apontando para uma lista de *sítes* com vídeos culinários, em que o usuário possivelmente encontrará outros vídeos sobre o tema, caso os recursos oferecidos pelo dicionário não sejam suficientes para esclarecer suas dúvidas.

Quando apropriado, serão apresentados sinônimos, tanto para a entrada em LP (“**UTES concorrentes em LP**”) quanto para a forma equivalente em LCh, listadas logo após a forma equivalente (“**Tradução**”), por ordem decrescente de frequência no corpus. A propósito, todos os sinônimos oferecidos no verbete serão acompanhados de suas respectivas frequências no corpus (“**freq. no corpus**”). No caso dos sinônimos em LP, será apresentado um link para a página de buscas do Google, para que o usuário possa conferir a frequência da UTE na *web*. O verbete apresentará também uma lista de outras entradas relacionadas em LCh, seja por conterem um ou mais dos itens lexicais da UTE-entrada, ou uma categoria que englobe algum desses itens, p. ex.: o verbete de “extra-virgin olive oil” apresentará, nos *links* para “**UTES relacionadas**”, tanto “azeite-de-dendê” quanto “regue com um fio de [ÓLEO]”, etc.

Por fim, assim que for possível desenvolver a representação visual da Culinária de forma adequada, a entrada, quando apropriado, conterà *links* apontando para os locais em que a UTEs ocorre na organização daquele gênero, tipologia textual ou parte do texto, p. ex.: o verbete de “enquanto isso” vai apontar para a porção do tipo textual “receita” que trata do “modo de preparo”.

5.1.3.3 Paradigma pragmático

Este é outro grande diferencial do verbete do DOC-Trad pois apresenta, além de 3 exemplos originalmente produzidos em LP com respectivas traduções em LCh, dois links em que o usuário pode ainda acessar até 10 outros contextos bilíngües (provenientes do corpus paralelo de culinária do CorTrad) ou até 10 contextos monolíngües contendo o padrão (ou padrões) equivalentes na língua de chegada, que são exibidos numa janela, na própria página, logo abaixo do link acionado.

Caso o usuário não consiga dirimir suas dúvidas por meio da consulta a esses contextos de uso, pode então acessar as versões integrais dos corpora paralelo (**Consultar CorTrad**) e comparável (**Consultar CorTec**) por meio do link no menu lateral à esquerda. Pode, também, usar o **KWIC Google** para examinar ocorrências obtidas no *Google Seach*, bem como consultar outras fontes de pesquisa da área e diversos motores de busca, listados no link “**Outras fontes de pesquisa**”.

5.1.3.4 Paradigma de forma equivalente

No Capítulo II (item 2.4) apresentamos a categorização de Schmitt (1989 *apud* Azenha Jr., 1999: 79) para os problemas de equivalência entre as línguas, bem como as estratégias de tradução propostas por Barbosa (1990: 93) para os casos de não-equivalência e equivalência parcial. Por considerarmos o fenômeno da equivalência do ponto de vista textual, ou seja, interessa-nos saber como as equivalências, não-equivalências e equivalências parciais poderiam ser representadas lingüística e textualmente em receitas traduzidas, para efeitos deste dicionário, trabalhamos com as seguintes definições:

- **equivalência:** casos em que **é possível identificar, na língua de chegada, um equivalente consagrado pelo uso** para uma determinada unidade de tradução especializada do texto de partida, quer ela se situe no nível da palavra, do sintagma, da oração, da organização textual, da cultura, etc. Nesses casos, não haveria necessidade de se adicionar qualquer comentário para o tradutor no verbete do dicionário;
- **equivalência parcial:** casos em que **uma unidade de tradução especializada consagrada pelo uso** na língua de chegada **precisa sofrer acréscimos ou omissões** para que possa cumprir a função de equivalente de uma unidade de tradução do texto de partida. Neste caso, a UTE oferecida no campo “**Tradução**” conterá alguma característica visual ou marca tipográfica que a distinga dos demais equivalentes, e o campo “**Comentários para o tradutor**” trará a explicação das diferenças e ajustes possíveis / necessários para usar a forma equivalente nos contextos mais recorrentes;
- **não-equivalência:** casos em que **não é possível identificar, na língua de chegada, um equivalente consagrado pelo uso** para uma determinada unidade de tradução especializada do texto de partida, quer ela se situe no nível da palavra, do sintagma, da oração, da organização textual, da cultura, etc. Nesses casos, o campo “**Tradução**” ficará em branco (ou conterá a expressão “*vide comentários*”), e o campo “**Comentários para o tradutor**” incluirá uma explicação que possa ser incorporada numa nota de tradução, acompanhada de sugestões de substitutos (que, quando não puderem ser identificados nos materiais de apoio disponíveis, deverão ser pesquisados

junto a especialistas da área que tenham um bom conhecimento das duas culturas em questão).

5.1.4 Exemplos de verbete

Os exemplos das figuras 77 e 78 mostram a aparência da página de consulta para um verbete na direção português-inglês (“pimenta-do-reino moída na hora”) e para um verbete na direção inglês-português (“[FOOD], sliced”) quando o link que dá acesso às linhas de concordância na língua de chegada é acionado pelo usuário (um exemplo da página inicial do verbete, sem esse link ativado, pode ser visto no Anexo VII).

5.1.4.1 Direção português-inglês

STU (Specialized Translation Unit):

pimenta-do-reino moída na hora

freq. SL corpus:

213

Translation:

freshly ground black pepper

freshly ground pepper

freq. TL corpus:

2.446

249

[our files](#)
[web videos](#)

[other sources: Google Image](#)

Translator's notes:

Although in English the adjective "black" is seldom omitted, in Portuguese, the corresponding adjective, "preta", is used mostly when there are other types of pepper being used in the recipe.

Examples (original Portuguese → translated English):

Esfregue o pernil com azeite, sal e pimenta-do-reino moída na hora	Rub fresh ham with olive oil, salt and freshly ground black pepper
Sal e pimenta-do-reino (moída na hora, de preferência) a gosto	Salt and black pepper (preferably freshly ground), to taste
Tempere os camarões com o sal e a pimenta moída na hora	Season shrimps with salt and freshly ground pepper

→ [show more bilingual contexts from CorTrad parallel corpus](#) (up to 10)

→ [show more contexts from CorTec comparable corpus – English](#) (up to 10)

ge handful of mint, chopped salt and **freshly ground black pepper** Method 1. To make the chi
thyme olive oil flaked sea salt and **freshly ground black pepper** 250ml / 8fl oz chicken stock...
cluding the oil). Season with salt and **freshly ground black pepper** and stir well. 3. Cover and chill
lb garlic, cut in half coarse salt and **freshly ground black pepper** extra virgin olive oil To serve:
season the beefsteaks with salt and **freshly ground black pepper**. 6. Heat a griddle pan until se
cream 1 egg yolk pinch of salt and **freshly ground black pepper** 110g / 4oz spinach, chopped
and season as you like with salt and **freshly ground black pepper**. Serve at once.
the flour with plenty of salt and **freshly ground pepper**. 5. Coat the rings in flour and
s coarsely chopped dill 1/2 teaspoon **freshly ground black pepper** 1/3 cup sour cream 2 teaspo
heat. 3. Season the marlin with **freshly ground pepper** then coat the top of the steak

Concurrent STUs in SL:

pimenta moída na hora

pimenta-do-reino preta moída na hora

freq. in corpus:

6

6

Related STUs in TL:

pepper₁; black pepper; white pepper; pink pepper; green peppercorns; salt and pepper to taste; mixed
peppercorns; coarsely ground PEPPER₁; pepper shaker; pepper mill; sprinkle with salt and pepper;
freshly ground PEPPER₁; cracked PEPPER₁

links to the hypertree:

[ingredient](#)
[process](#)

Figura 77: exemplo de verbete português-inglês

5.1.4.2 Direção inglês-português

<p>UTE (Unidade de Tradução Especializada):</p> <div style="border: 1px solid black; padding: 2px; margin-bottom: 5px;">[FOOD], sliced</div> <p>Tradução:</p> <div style="border: 1px solid black; padding: 2px; margin-bottom: 5px;">[ALIMENTO] (cortado) em fatias</div> <div style="border: 1px solid black; padding: 2px; margin-bottom: 5px;">[ALIMENTO] (cortado) em rodela(s)</div> <div style="border: 1px solid black; padding: 2px;">[ALIMENTO] fatiado</div>	<p>freq. corpus LP:</p> <div style="border: 1px solid black; padding: 2px; margin-bottom: 5px;">2.820</div> <p>freq. corpus LCh:</p> <div style="border: 1px solid black; padding: 2px; margin-bottom: 5px;">1.245</div> <div style="border: 1px solid black; padding: 2px; margin-bottom: 5px;">935</div> <div style="border: 1px solid black; padding: 2px;">377</div>	<p>nosso arquivo vídeos da web outras fontes: Google Image</p>
---	---	---

Comentários para o tradutor:

Ocorre tipicamente na lista de ingredientes. ALIMENTOS que mais co-ocorrem com “(cortado) em fatias”: pão (613 oc.), presunto (246), queijo (215), bacon (206), cebola(s) (171), mussarela (150), alho-poró (98); ALIMENTOS que mais co-ocorrem com “(cortado) em rodela(s)”: cebola(s) (344 oc.), tomate(s) (209), limão (112), batata(s) (99), piment(ão/ões) (80); ALIMENTOS que mais co-ocorrem com “fatiado”: cebola (89 oc.), presunto (46), mussarela (32), cogumelo (24).

Exemplos (original English → translated Portuguese):

3 small cloves garlic , thinly sliced lengthwise	3 dentes de alho pequenos, cortados em fatias finas no sentido do comprimento
2 yellow peppers , sliced	2 pimentões amarelos em rodela(s)
55g / 2oz cheddar cheese , sliced	55g de queijo cheddar fatiado

→ [mostrar mais contextos bilíngües no corpus paralelo CorTrad](#) (até 10 ocorrências)

→ [mostrar contextos no corpus comparável CorTec – inglês](#) (até 10 ocorrências)

ainha 4 col. (sopa) de azeite de oliva 1 **limão cortado em rodela(s)** suco de ½ limão 1 col. (chá) de alho picado sal alho-poró Ingredientes 2 talos de **alho-poró cortado em fatias** (só a parte branca) 20 g de manteiga 2 xíc. (chá) tiras finas 1 dente de alho inteiro 1 **cebola grande, em rodela(s)** Sal, pimenta-do-reino e alecrim a gosto Prepa 1 cebola cortada em rodela(s) 1 **cenoura cortada em rodela(s)** sal e pimenta-do-reino Farofa 1/2 con rodela(s) 2 pimentões verdes ou vermelhos **cortados em rodela(s)** Sal e pimenta-do-reino a gosto Preparo 1. T nha cozidas e escorridas 300 g de **queijo mussarela em fatias** Para polvilhar Farinha de rosca Queijo parmes de manteiga 300 g de **palmito em conserva cortado em rodela(s)** Para polvilhar 1 colher (chá) de manjerição p cortadas em tiras 2 dentes de alho picados 1 **cebola fatiada** Para pincelar 1 ovo batido Acessório forma /2 colher (chá) de fermento em pó 2 **cebolinhas médias em rodela(s) grossas** Sal a gosto Para fritar e untar Óleo 1 pimentão amarelo sem pele e sementes 1 **berinjela em fatias finas** 1 abobrinha em fatias finas salsa picada

<p>UTES concorrentes na LP:</p> <div style="border: 1px solid black; padding: 2px; margin-bottom: 5px;">[FOOD], cut into rings</div> <p>UTES relacionadas:</p> <div style="border: 1px solid black; padding: 5px; margin-top: 5px;">fatia(s) de [ALIMENTO]; rodela(s) de [ALIMENTO]; anéis de [ALIMENTO]; [ALIMENTO] (cortado) em anéis; [ALIMENTO] (cortado) em lâminas; [ALIMENTO] laminado</div>	<p>freq. no corpus:</p> <div style="border: 1px solid black; padding: 2px; margin-bottom: 5px;">14</div>
---	--

Links para árvore hiperbólica: [processo](#)

Figura 78: Exemplo de verbete português-inglês

This document was created with Win2PDF available at <http://www.daneprairie.com>.
The unregistered version of Win2PDF is for evaluation or non-commercial use only.

Conclusão

Ao longo deste trabalho, procurei apresentar evidências que comprovassem minha tese de que um dicionário técnico, para que seja mais útil e confiável para o tradutor, deve apresentar informações não apenas relativas aos termos e conceitos de uma dada área especializada (o que tem sido feito, com maior ou menor êxito, usando-se as metodologias de trabalho da Terminologia), mas também informações relativas à padronização textual. Isto é, deve registrar as regularidades lingüísticas que tipificam esses textos, ainda que as mesmas compreendam fenômenos que, por não estarem diretamente relacionados com a terminologia da área *stricto sensu*, têm sido vistos apenas como manifestações da “língua geral” e, portanto, desconsiderados pelas obras terminográficas.

Parti do pressuposto de que o tradutor não segmenta o texto em “termos” e “conceitos” ao traduzir, mas em unidades textuais (ou unidades de tradução) vinculadas a situações de comunicação específicas que, no texto técnico, convencionei chamar de Unidades de Tradução Especializadas (UTES). É para essas UTES que o tradutor precisa de equivalentes, independente de serem / conterem termos ou não.

São elas, portanto, que também devem constar dos dicionários técnicos voltados para este público.

Ora, para compilar essas unidades, era preciso uma abordagem diferenciada da que vem sendo utilizada na elaboração da maioria dos dicionários bi- e multilíngües técnicos disponíveis atualmente no Brasil. Uma abordagem que possibilitasse ver o texto técnico não mais como um repositório de termos e definições, mas como um sistema de probabilidades em que certas UTEs, de acordo com a cultura, o gênero, o tipo textual e o contexto de uso, são mais típicas (i.e., recorrentes) do que outras.

A Lingüística de Corpus, que tem se dedicado à identificação de padrões léxico-gramaticais recorrentes por meio da observação empírica de textos autênticos organizados em *corpora* eletrônicos, foi a abordagem teórico-metodológica eleita para levar a cabo essa tarefa. Seus conceitos básicos e algumas das principais ferramentas de exploração de *corpora* de que dispõe foram apresentados no Capítulo III, que também descreveu como foi compilado o Corpus Comparável de Culinária, com cerca de 1,5 milhão de palavras em cada língua, que serve de base para a pesquisa. A Culinária foi privilegiada como área técnica neste estudo, pela demanda crescente por traduções e versões que tem gerado no Brasil nos últimos anos. As características específicas da tradução técnica nessa área foram discutidas no Capítulo II.

O objetivo último deste trabalho era propor um Dicionário *Online* de Culinária para o tradutor de Culinária que abrigasse as UTEs coletadas. Mas uma parte significativa da pesquisa prática consistiu em descrever e testar metodologias "manuais" (que partem da observação de linhas de concordância) e semi-automáticas (que fazem uso de programação e coleta em massa de dados) em Lingüística de Corpus para a identificação dessas UTEs em *corpora* especializados.

Os resultados obtidos na primeira etapa dessa pesquisa prática – análise "manual" das ocorrências de "flour*" em uma porção do *corpus* de inglês – confirmaram que as UTEs recorrentes em textos especializados da área da Culinária extrapolam o que se costuma considerar como unidades terminológicas. Além disso (e talvez por isso), como demonstrado no caso dos padrões relacionados a "flour*", em sua grande maioria essas UTEs são desprezadas pelas obras de referência bilíngües e monolíngües dessa área disponíveis no Brasil atualmente. Isso aponta para a necessidade de que sejam usadas metodologias de compilação de dicionários técnicos

que abordem o texto técnico de uma outra perspectiva, a fim de oferecer ao tradutor as informações de que necessita para a produção textual na língua de chegada.

Como a metodologia empregada no levantamento “manual” de UTEs em parte do *corpus* de inglês mostrou-se bastante demorada e trabalhosa, a segunda parte do capítulo dedicou-se a descrever algumas metodologias de extração semi-automática de UTEs, que foram testadas, desta vez, na versão completa do *corpus* de português. Num primeiro momento, na tentativa de caracterizar as UTEs da Culinária quanto aos seus padrões morfossintáticos e semânticos, foram analisadas cerca de 1.300 UTEs provenientes do glossário que compilei em minha memória de tradução da área da Culinária para saber: i) a categoria gramatical de cada item lexical que compunha as UTEs plurivocabulares e de cada UTE como um todo; ii) a função sintática das UTEs; iii) se as UTEs continham ou não palavras-chave; iv) a frequência das UTEs no *corpus* total. Em seguida, os padrões morfossintáticos identificados nessa primeira fase foram usados para extrair padrões semelhantes de forma semi-automática na versão etiquetada do *corpus* de português.

De modo geral, os dados obtidos apontam para a relevância de se trabalhar com um *corpus* cuja etiquetagem tenha sido revisada para se conformar às particularidades léxico-gramaticais da área especializada em questão, já que o sucesso das metodologias de extração semi-automáticas de base lingüística depende, em grande medida, da acuidade das etiquetas atribuídas aos itens lexicais do corpus. Por outro lado, concluí que a extração “manual” de UTEs a partir da lista de palavras-chave do corpus, ainda que seja lenta e custosa, é a que produz melhores resultados, em se considerando que a extração automática e semi-automática eficientes, até o momento, dependem de um conhecimento computacional mais aprofundado por parte do pesquisador – o que, muitas vezes, pode ter um custo mais alto em termos de tempo e dificuldade de execução do que teria o exame “manual” de todas as linhas de concordância geradas para cada palavra-chave do corpus. Isso aponta para a importância de se estabelecerem parcerias entre os pesquisadores da área de Letras e os lingüistas computacionais que trabalham no âmbito do Processamento da Linguagem Natural. Por fim, foi observado que as UTEs presentes nos textos técnicos da Culinária, além de apresentarem padrões muito mais longos do que se costuma observar nas entradas de dicionários técnicos, não se restringem aos substantivos e verbos – os advérbios e as preposições, por exemplo, têm presença marcante nos

padrões da área e, por serem fundamentais para a produção textual nessa área, devem constar também de um dicionário voltado para o tradutor.

Para abrigar as UTEs, elaborei uma proposta de Dicionário cuja principal característica é apresentar vários recursos que auxiliam o tradutor na etapa da tradução que envolve a produção textual na língua de chegada, tais como: comentários sobre o grau de equivalência parcial ou não equivalência entre UTEs nas duas línguas / culturas; pelo menos 3 exemplos originalmente produzidos na língua de partida, acompanhados de suas respectivas traduções na língua de chegada, e a possibilidade de exibir, na própria página do verbete, até 10 exemplos de uso da UTE equivalente na língua de chegada, e até 10 exemplos paralelos, com original e tradução. Caso não sejam suficientes, o tradutor pode ainda consultar a versão integral dos *corpora* comparável e paralelo (via *CorTec* e *CorTrad*, respectivamente), ou gerar linhas de concordância a partir de resultados gerados pelo Google, usando a ferramenta *KWIC Google*. Caso a dúvida seja conceitual, o tradutor pode pesquisar os vários *links* para *sites* sobre culinária em inglês e português oferecidos pela interface *online* do dicionário, ou até mesmo postar sua dúvida no Fórum, para receber ajuda de outros internautas – informação que poderá ser usada por nós para identificar lacunas no dicionário e, então, preenchê-las.

Espero, com isso, ter demonstrado que a **Linguística de Corpus**, se usada não apenas como metodologia, mas como **abordagem teórica** na exploração de *corpora* especializados, contribui para a elaboração de obras de referência mais úteis e confiáveis para o tradutor, pois a **visão de língua** como sistema probabilístico que a sustenta permite considerar como entradas relevantes quaisquer padrões associativos com uma frequência significativa em *corpora* de textos representativos da área. Isto é, a **probabilidade** parece ser um critério mais relevante para a seleção do repertório de entradas de um dicionário técnico voltado para o tradutor-produtor-de-textos do que sua pertinência a um esquema estruturado de conhecimento que, em tese, representaria uma dada área de conhecimento. Daí minha proposta de se partir de listas de palavras-chave de *corpora* representativos e dos padrões associados a essas palavras para definir a lista de entradas de um dicionário voltado para o tradutor.

A abordagem proposta aqui, se utilizada na análise de *corpora* de outras tipologias textuais da Culinária, ou na compilação de dicionários técnicos de outras áreas especializadas, poderia suscitar a necessidade de ajustes na metodologia e/ou no modelo de verbete para que fossem levadas em conta as características relevantes

para as diferentes áreas, gêneros e tipologias textuais. Isso aponta para a necessidade de se pesquisar, em trabalhos futuros, outras formas de transpor para modelos terminológicos e terminográficos essa visão de língua como sistema probabilístico, bem como as necessidades específicas do tradutor-produtor-de-textos nas diversas áreas do saber. Essas pesquisas poderiam, quiçá, instaurar uma nova área de estudos aplicados, situada na confluência da Terminografia com a Tradução técnica e a Lingüística de Corpus – Terminologia Tradutiva, a meu ver, seria um nome possível para essa área.

Em suma, é preciso que os dicionários técnicos bi- e multilíngües passem a considerar o fato de que o tradutor técnico não traduz termos, mas unidades de tradução, que estão inseridas em textos, produzidos no âmbito de uma tipologia textual, de um gênero, de uma cultura. Um dicionário que seja útil ao tradutor na produção textual em língua de chegada deve registrar essas unidades, sejam elas termos ou não.

This document was created with Win2PDF available at <http://www.daneprairie.com>.
The unregistered version of Win2PDF is for evaluation or non-commercial use only.

Referências bibliográficas

ABBADE, C.M.S. (2005). Receitas culinárias medievais. *Cadernos do CNLF, Volume IX, no.14*. Disponível em: <http://www.filologia.org.br/ixcnlf/14/12.htm> [acesso em: 21/08/08].

AIJMER, K. (1996). *Conversational routines in English. Convention and creativity*. London/New York: Longman.

ALAMBERT, E.G.S. (2008). *Uma tradução premiada sob a perspectiva da lingüística de corpus*. Dissertação (Mestrado em Lingüística Aplicada e Estudos da Linguagem) – Pontifícia Universidade Católica de São Paulo, São Paulo. Disponível em: http://www.pucsp.br/pos/lael/lael-inf/teses/eliane_alambert.pdf [acesso em 10/10/08]

ALGRANTI, M. (2000). *Pequeno Dicionário da Gula*. Rio de Janeiro: Record.

ALLEN, G. (1999). *Resource guide for food writers*. New York: Routledge.

ALTENBERG, B. e S. GRANGER (eds.). (2002). *Lexis in Contrast: Corpus-based approaches*. Amsterdam: John Benjamins.

ALTENBERG, B. & M. EEG-OLOFSSON (1990). Phraseology in spoken English: Presentation of a project. In AARTS, J. & W. MEIJS (eds.) *Theory and Practice in Corpus Linguistics*. Amsterdam: Rodopi, p. 1–27.

ALVES, F.; C. MAGALHÃES e A. PAGANO (2003 [2000]). *Traduzir com autonomia: Estratégias para o tradutor em formação*. São Paulo: Contexto, 2000.

ALVES, I.C. (1983) *Modalidades de tradução: uma avaliação do modelo proposto por Vinay e Darbelnet*. Dissertação de mestrado. São Paulo, PUCSP.

ALVES, I.M. (org.). (2007). *Ciências do léxico: lexicologia, lexicografia, terminologia*. 1 ed. Campo Grande/São Paulo: Editora da UFMS/Humanitas, 2007, v. III.

_____. (2006). A renovação lexical nos domínios de especialidade. *Revista Ciência e Cultura* vol.58, no.2 (abril/junho). São Paulo, p. 48-51. Disponível em: http://cienciaecultura.bvs.br/scielo.php?pid=S0009-67252006000200013&script=sci_arttext [acesso em: 30/07/06].

AMMAN, M. (1989). Landeskunde in der Translationsausübung. *TextconText* 4, p. 90- 105.

AMMAN, M. e H.J. VERMEER (1990). Entwurf eines Curriculums für einen Studiengang Translatologie und Translatorik. Heidelberg.

ARAÚJO, W.M.C., A.P.R. PERETTI, C.M.R. TENSER. N. PILLA e M.C. FLAUSINO (orgs.) (2004). *Coletânea de Palestras do 1o. Encontro Brasileiro de Gastronomia e Segurança Alimentar*. Brasília: UnB.

ARRIBAS BAÑO, M.A. (2005). *La representación del significado en el diccionario bilingüe de negocios: una propuesta de tratamiento*. Tese (Doctorado en Filología Inglesa). Facultad de Filosofía y Letras de la Universidad de Valladolid.

ATKINS, B.T.S. (org.) (1998). *Using dictionaries: studies of dictionary use by language learners and translators*. Tübingen: Niemeyer.

ATKINS B.T.S. e M. RUNDELL (2008). *The Oxford Guide to Practical Lexicography*. United Kingdom: Oxford University Press.

AUBERT, F. H. (2003a). Traduzindo as diferenças extralingüísticas – procedimentos e condicionantes. *TradTerm*, São Paulo, v. 9, p. 151-172.

_____. (2003b). As variedades de empréstimos. *DELTA*, São Paulo, v. 19, p. 27-42.

_____. (2001). Língua como estrutura e como fato histórico-social: conseqüências para a terminologia. In: ALVES, I.M. (org). *A constituição da normalização terminológica no Brasil*. 2a. ed. São Paulo: Humanitas FFLCH/USP, p. 11-15.

_____. (1996). *Introdução à metodologia da pesquisa terminológica bilíngüe*. São Paulo: Humanitas FFLCH/USP.

_____. (1992). Problemas e urgências na interrelação terminologia / tradução. *Alfa*, v. 36, São Paulo, p. 81-86.

AUGER, P. (1988). La Terminologie au Québec et dans le monde de la naissance à la maturité. *Actes du Sixième Colloque OLF-STQ de Terminologie*. 1985. Québec: Office de la Langue Française et Société des Traducteurs du Québec, p. 27-59.

_____. (1984). La Commission de Terminologie de l'Office de la Langue Française et la Normalisation Terminologique. *Terminogramme*, 26: 9-12.

AUGER, P. e M.-C. L'HOMME (2004 [1994-5]). A terminologia segundo uma abordagem textual: uma representação mais adequada do léxico nas linguagens de especialidade. [tradução: Cristina de Campos Velho Birck]. In: KRIEGER. M.G. e L. ARAÚJO (orgs.) (2004).

- AVANCE DE SOUZA, E. (2003). *Colocações verbais no âmbito dos contratos sociais e estatutos: estudo contrastivo alemão/português*. Dissertação (Mestrado em Língua e Literatura Alemã). Faculdade de Filosofia, Letras e Ciências Humanas da USP, São Paulo.
- AZENHA JR., J. (1999). *Tradução técnica e condicionantes culturais. Primeiros passos para um estudo integrado*. São Paulo: Humanitas FFLCH/USP.
- BAAKES, K. (1994). *Key issues of syntax in the special languages of science and technology*. Heidelberg: Julius Groos Verlag.
- BACCIN, P. G. (2003). *Italianismos na terminologia gastronômica na cidade de São Paulo: um modelo de glossário*. Tese (Doutorado em Filologia e Língua Portuguesa) – Faculdade de Filosofia, Letras e Ciências Humanas da Universidade de São Paulo, São Paulo, 2v.
- BAKER, M. (ed.) (1998). *Routledge Encyclopedia of Translation Studies*. London: Routledge.
- _____. (1995). Corpora in Translation Studies: An Overview and Some Suggestions for Future Research. *Amsterdã, Target 7*, vol. 2, p. 223-243.
- _____. (1993). Corpus linguistics and translation studies. Implications and applications. In: BAKER, M.; G. FRANCIS e E. TOGNINI-BONELLI (eds.). *Text and technology. In honour of John Sinclair*. Amsterdam: John Benjamins, pp. 233-250.
- _____. (1992). *In other words: a coursebook on translation*. London and New York: Routledge.
- BARBOSA, H.G. (1990). *Procedimentos Técnicos da Tradução: uma nova proposta*. Campinas: Pontes.
- BARBOSA, M.A. (2006). Para uma Etno-terminologia: recortes epistemológicos. *Revista Ciência e Cultura* vol.58, no.2 (abril/junho). São Paulo, p. 48-51. Disponível em: http://cienciaecultura.bvs.br/scielo.php?pid=S0009-67252006000200018&script=sci_arttext [acesso em: 30/07/06].
- _____. (1998). Terminologização, vocabularização, cientificidade, banalização: relações. *Acta semiótica et linguística*, São Paulo, v. 7, p. 25-44.
- BARCELLOS, G. A. *Ferramentas Computacionais na Pesquisa Terminológica*. (Palestra na Jornada de Tradução do CITRAT - junho de 2006).
- BARLOW, M. e S. KEMMER. (1994). A Schema-based Approach to Grammatical Description. In S. LIMA, R. CORRIGAN e G. IVERSON (eds.). *The Reality of Linguistic Rules*, Amsterdam: Benjamins.
- BARONE, V (1999). Para ler e degustar. *Revista Época*, edição 38 (08/02/1999). Disponível em: <http://epoca.globo.com/edic/19990208/socied4.htm> [acesso em: 24/04/2008].
- BARRETO, R. (2002). *Passaporte para o sabor: Tecnologias para a elaboração de cardápios*. Editora Senac São Paulo.
- BARROS, L.A. (2007). *Conhecimentos de terminologia geral para a prática tradutória*. São José do Rio Preto: NovaGraf.

_____. (2006). Aspectos epistemológicos e perspectivas científicas da terminologia. *Revista Ciência e Cultura* vol.58, no.2 (abril/junho). São Paulo, p. 22-26. Disponível em: http://cienciaecultura.bvs.br/scielo.php?pid=S0009-67252006000200011&script=sci_arttext [acesso em: 30/07/06].

_____. (2004). *Curso básico de terminologia*. São Paulo: Edusp.

BASSNETT-McGUIRE, S. (1980). *Translation studies*. London: Methuen & Co. Ltd.

BEAUGRANDE, R. (1978). *Factors in a Theory of Poetic Translating*. Assen, The Netherlands: Van Gorcum.

BELL, R.T. (1991). *Translation and translating*. London: Longman.

BERBER SARDINHA, A.P. (2008). Metaphor probabilities in corpora. In: ZANOTTO, M.S.; L. CAMERON e M.C. CAVALCANTI (eds.). *Confronting Metaphor in Use: An applied linguistic approach*. Amsterdam / Philadelphia: John Benjamins, pp. 127-148.

_____. (2004). *Lingüística de Corpus*. Barueri: Manole, 2004.

_____. (2000). Prosódia Semântica na Tradução do Português e Inglês: um Estudo Baseado em Corpus". In: VOLPE NUNES, M.G. (ed.). *Anais do V Encontro para o processamento computacional da língua portuguesa escrita e falada (PROPOR 2000)* (Atibaia, SP, 19-22 de Novembro de 2000), São Paulo: ICMC/USP, pp. 93-104.

_____. (2002). Corpora eletrônicos na pesquisa em tradução. In: TAGNIN (org.) (2002c).

_____. (1998). Size of a representative corpus. *Resumo de discussão sobre CORPORA*, lista de discussão, 26/08/1998. Disponível em: <http://nora.hd.uib.no/corpora/1998-3/0120.html> [acesso em 10/10/08]

BERNARDINI, S. e F. ZANETTIN (eds.) (2000). *I corpora nella didattica della traduzione. Corpus Use and Learning to Translate*. Bologna: CLUEB.

BEVILACQUA, C.R. (2004). *Unidades Fraseológicas Especializadas Eventivas: descripción y reglas de formación en el ámbito de la energía solar*. Tese (Doutorado em Lingüística Aplicada), Instituto Universitário de Lingüística Aplicada, Universidade Pompeu Fabra, Barcelona.

_____. (2001). Unidades fraseológicas especializadas: novas perspectivas para sua identificação e tratamento. In: KRIEGER, M.G. e A.M.B. MACIEL (orgs). *Temas de Terminologia*. São Paulo / Porto Alegre: Humanitas FFLCH/USP / Editora da Universidade, p. 106-117.

_____. (1999). *Unidades Fraseológicas Especializadas: estado de la cuestión y perspectivas*. Barcelona: Universitat Pompeu Fabra, Institut Universitari de Lingüística Aplicada. Trabalho de pesquisa.

_____. (1996). *A fraseologia jurídico-ambiental*. Dissertação (Mestrado em Estudos da Linguagem) – Departamento de Letras da Universidade Federal do Rio Grande do Sul, Porto Alegre.

BIBER, D. (1993). Representativeness in corpus design. *Literary and Linguistic Computing* 8, 4, p. 243-257.

_____ (1990). Methodological issues regarding corpus-based analyses of linguistic variation. *Literary and Linguistic Computing* 5, p. 257-269.

BIBER, D.; S. CONRAD e R. REPPEN (1998). *Corpus Linguistics. Investigating language structure and use*. Cambridge: CUP.

BIDERMAN, M.T.C. (2004). Análise de dois dicionários gerais do português contemporâneo: o Aurélio e o Houaiss. In: ISQUERDO, A.N. e KRIEGER, M.G. *As ciências do léxico. Volume III*. Campo Grande: Editora UFMS, 2004, p.185-200.

BLAIS, E. (1993). La phraséologie : une hypothèse de travail. *Terminologies Nouvelles*, Bruxelles, n. 10, p. 50-56.

BOURIGAULT, D. e M. SLODZIAN (2004 [1999]). Por uma terminologia textual. [tradução: Sandra Dias Loguercio]. In: KRIEGER e ARAÚJO (orgs.) (2004).

BOWKER, L. (2000). A Corpus-Based Approach to Evaluating Student Translations. *The Translator*. Volume 6, number 2. Manchester: St Jerome Publishing.

_____ (1996). Towards a Corpus-Based Approach to Terminography. *Terminology*, 3(1), 27-52.

BOWKER, L. e J. PEARSON (2002). *Working with Specialized Language*. London and New York: Routledge.

BYRNE, J. (2006). *Technical Translation: Usability Strategies for Translating Technical Documentation*. The Netherlands: Springer.

CABRÉ Castelvì, M.T. (2005). La terminología y la terminografía hoy: concepción, proceso y herramientas. *Curso ministrado durante o XVII Instituto Brasileiro de Lingüística*. Brasília: UnB (carga horária: 15 horas).

_____ (2003). Theories of terminology. *Terminology* 9:2, p. 163-199.

_____ (2002). Análisis textual y terminología, factores de activación de la competencia cognitiva en la traducción. In: ALCINA CAUDET, A. e S. GAMARO PÉREZ (eds). *La traducción científico-técnica y la terminología en la sociedad de la información*. Castellón: Publicacions de la Universitat Jaume I, p. 87-105.

_____ (1999a). *Terminología: Representación y Comunicación: elementos para una teoría de base comunicativa y otros artículos*. Barcelona: Universitat Pompeu Fabra. Sèrie Monografies, 3.

_____ (1999b). El Traductor y la terminología. *Colloquio internacional: interpretar, traducir textos de la(s) cultura(s) hispánica(s)*. Università di Bologna. Forlì 21-23 de oct.

CABRÉ, M.T. LORENTE, M. e ESTOPÁ, R. (1996). Terminología e fraseología. *Actas del V Simposio de Terminología Iberoamericana*. Ciudad de México: RITERM, nov. de 1996, p. 67-81.

CARLI, F. e E. KLOTZ (2007). *Dicionário gastronômico*. São Paulo: Contorno Editora.

CARNEIRO, (2004). Comida e sociedade: significados sociais na história da alimentação. In: ARAÚJO *et al.* (org.) (2004).

CARVALHO, L. (2007). *A tradução de binômios em contratos de common law à luz da lingüística de corpus*. Dissertação (Mestrado em Estudos Lingüísticos e Literários em Inglês) – Departamento de Letras Modernas da Faculdade de Filosofia, Letras e Ciências Humanas da Universidade de São Paulo, São Paulo, 2v.

CASTANHO, R.M.C. (2003). *Proposta para elaboração de um glossário de colocações na área médica – subárea hipertensão arterial*. Dissertação (Mestrado em Estudos Lingüísticos e Literários em Inglês) - Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo.

CATFORD, J.C. (1965). *A Linguistic Theory of Translation: An Essay in Applied Linguistics*. Oxford: Oxford University Press.

CHAMPOLLION, Y. (2001). *WordFast: Translation Memory Engine*. Disponível em: <http://www.wordfast.org/> [acesso em 10/10/08]

CHAPELLE, C. A. (1998) Some notes on Systemic-Functional linguistics. Artigo eletrônico do website *Systemic Functional Linguistics Information Site*. Disponível em: <http://www.isfla.org/Systemics/Definition/chapelle.html> [acesso em 19/09/08].

CHESTERMAN, A. (1998). *Contrastive functional analysis*. Amsterdam: John Benjamins.

CHOMSKY, N. (1957) *Syntactic structures*. The Hague: Mouton.

CHURCH, K.W. e W.A. GALE (1991). Concordances for parallel texts. *Using corpora. Proceedings of the Seventh Annual Conference of the UW Centre for the New OED and Text Research* (Oxford, 29 Sept-1 Oct 1991), p. 40-62.

CIVITA, V. (org). (1974). *ABC da Cozinha*. São Paulo: Abril Cultural.

COINTREAU, E. (2005). *10 years of cookbooks*. Press release do Gourmand World Cookbook Award. Disponível em: http://www.cookbookfair.com/html/10_years_of_cookbook.html [acesso em: 19/08/08].

COLINA, S. (1997). Contrastive Rhetoric and Text-Typological Conventions in Translation Teaching. Amsterdam, *Target* 9, p. 335-353.

COMFORT, J. (1997). *Writing cookbooks*. Washington: Self-counsel Press.

COSTA, A.T.P. (2006). *Brasil mostrando a sua cara: estratégias de tradução no material de divulgação cultural – um estudo baseado em corpus*. Dissertação (Mestrado em Lingüística Aplicada) – Departamento de Línguas Estrangeiras e Tradução da Universidade de Brasília, Brasília.

COULTHARD, R. J. (2005). *The application of Corpus Methodology to Translation: the JPED parallel corpus and the Pediatrics comparable corpus*. Dissertação (Mestrado em Estudos da Tradução) – Universidade Federal de Santa Catarina, 2005.

COUTO, C. *Arte de cozinha - alimentação e dietética em Portugal e no Brasil (séculos XVII-XIX)*. São Paulo: Editora Senac de São Paulo, 2007.

DANCETTE, J. (1998). Le potentiel du dictionnaire spécialisé bilingue électronique: viser la discursivité ou la formalisation des relations sémantiques. In: FONTENELLE, T. et al. (eds.). *Actes EURALEX '98*, Université de Liège, Liège, p. 387-396.

DANCETTE, J. e M.-C. L'HOMME (2002). The Gate to Knowledge in a Multilingual Specialized Dictionary : Using Lexical Functions for Taxonomic and Partitive Relations. *Proceedings of the 10th EURALEX International Congress*. Copenhagen, Center for Sprogteknology, II, p. 597-606.

DANCETTE, J. e C. RÉTHORÉ (2000). *Dictionnaire Analytique de la Distribution. Analytical Dictionary of Retailing*. Montréal: Les Presses de l'Université de Montréal.

DARIN, L. (1986). *Translation modalities in the comparison of English and Portuguese – Analysis of excerpts taken from C. Castañeda's novel "The Teachings of D. Juan" or "A Erva-do-Diabo"*. MA thesis. Exeter.

DESMET, I (1995-6). *Por une approche terminologique des sciences sociales et humaines. Les sciences sociales et humaines du travail en portugais et en français*. (Tesis doctoral) – Paris, Université Paris-Nord (Paris XIII).

_____ (2004). *Modelos de análise semântica e de tratamento de equivalência em dicionários*. São Paulo, FFLCH/USP, 2004 (Palestra proferida na USP, São Paulo em 24 mar. 2004).

DÓRIA, C.A. (2006). *Estrelas no céu da boca: Escritos sobre culinária e gastronomia*. São Paulo: Editora Senac São Paulo.

EEG-OLOFSSON, M. e B. ALTENBERG (1996). Recurrent word combinations in the London-Lund Corpus: Coverage and use of word-class tagging. In: PERCY, C.E.; C. MEYER e I. LANCASHIRE (eds.). *Synchronic Corpus Linguistics. Papers from ICAME16*. Amsterdam e Atlanta: Rodopi, p. 97-108.

ENGLER, A. (2006). *Tradutor gastronômico*. [S.l.]: Associação Brasileira de Bares e Restaurantes.

ERICSSON, K.A. e H.A. SIMON (1984). *Protocol analysis: Verbal reports as data*. Cambridge, MA: MIT Press.

ESTOPÀ, R. (1999). *Extracció de terminologia: elements per la construcció d'un SEACUSE (Sistema d'Extracció Automàtica de Candidats de Significació Especializada)*. (Tesis doctoral) – Barcelona, Universitat Pompeu Fabra, Institut Universitari de Lingüística Aplicada.

FAULSTICH, E. (2006). A socioterminologia na comunicação científica e técnica. *Revista Ciência e Cultura* vol.58, no.2 (abril/junho). São Paulo, p. 48-51. Disponível em: http://cienciaecultura.bvs.br/scielo.php?script=sci_arttext&pid=S0009-67252006000200012&lng=en&nrm=iso [acesso em: 30/07/06].

FELBER, H. (1984). *Terminology Manual*. Paris: Unesco – International Information Centre for Terminology (Infoterm).

FILLMORE, C.J. (1985). Frames and the Semantics of Understanding. *Quaderni di Semantica* 6, p. 222-255.

_____ (1979). Innocence: A Second Idealization for Linguistics. *Berkeley Linguistic Society* 5, p. 63-76.

_____ (1978). On the Organization of Semantic Information in the Lexicon. In: FARKAS, D. et al (ed.). *Papers from the Parasession on the Lexicon*. Chicago: Chicago Linguistic Society, p. 148-173.

FILLMORE, C.J. e ATKINS, B.T.S. (1992). Towards a frame-based lexicon: The semantics of RISK and its neighbors. In: LEHRER, A. e KITTAY, E.F. (eds). *Frames, Fields and Contrasts*. New Jersey: Lawrence Erlbaum Associates Hillsdale, p. 75-102.

FINATTO, M.J.B. (2007). Exploração terminológica com apoio informatizado: diálogos entre terminologia e lingüística de corpus. In: Lorente, M.; R. Estopà, J. Freixa, J. Martí, C. Tebé (ed.). *Estudis de lingüística i de lingüística aplicada en honor de M. Teresa Cabré Castellví. Vol. 2: De deixebles*. Barcelona: IULA.

_____ (2004). Terminologia e lingüística de corpus: da perspectiva enunciativa aos novos enfoques do texto técnico-científico. *Letras de Hoje*, v. 39, n. 4, Porto Alegre, p. 97-106.

_____ (2001). Conhecendo um software do tipo *parser*: instrumento auxiliar para a pesquisa terminológica e sua utilização pelo estudante de letras tendo em vista a elaboração de glossários especializados. In: LIMA, M.S. e P.R. CHITTONI (orgs). *Canadá e Brasil: questões de ensino e terminologia*. Porto Alegre: UFRGS, Núcleo de Estudos Canadenses, ABECAN, vol. 1, p. 55-70.

FINATTO, M. J. B. ; SIMIONI, S. (2007). A causalidade no texto de Química: efeitos da tradução sobre a estrutura coesiva. *Calidoscópio (UNISINOS)*, v. 5, p. 177-184.

FINATTO, M. J. B. ; ZILIO, L. ; FICHTNER, M. F. (2006). Resíduos e Abfälle: um reconhecimento terminológico para a busca de equivalências entre o português e o alemão. *Tradterm*, v. 12, p. 269-292.

FIRTH, J.R. (1957). *Papers in Linguistics: 1934-1951*. London: Oxford University Press.

FISHBACK, H. (1998). Prefácio de editor convidado. In: *Translation and Medicine*, American Translators Association Series, Vol. X, pp. 1-12.

FORNARI, C. *Dicionário-Almanaque de Comers & Bebés*. Rio de Janeiro: Editora Nova Fronteira, 2001.

FRANCIS, G.; HUNSTON, S. e E. MANNING (1996). *Collins COBUILD Grammar Patterns 1: Verbs*. London: Harper Collins.

_____ (1998). *Collins COBUILD Grammar Patterns 2: Nouns and adjectives*. London: Harper Collins.

FRANCO, A. *De caçador a Gourmet: uma história da gastronomia*. São Paulo: Editora SENAC, 2001.

FRANCO AIXELÁ, J. (2004). The Study of Technical and Scientific Translation: An Examination of its Historical Development. JoSTrans, Issue 1. Disponível em: http://www.jostrans.org/issue01/art_aixela.php [acesso em 09/02/04]

FROMM, G. (2008). *VoTec: a construção de vocabulários eletrônicos para aprendizes de tradução*. Tese (Doutorado em Estudos Lingüísticos e Literários em Inglês) – Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo.

FUENTES MORÁN, M. T. e J. GARCÍA PALACIOS (2002). Los diccionarios de especialidad y el traductor. In: GUERRERO Ramos, G. y M. F. PÉREZ Lagos (eds). *Panorama actual de la terminología*. Málaga: Comares, p. 117-136.

FUNDAÇÃO ALÍCIA E ELBULLITALLER (2007). *Léxico científico-gastronômico: as chaves para entender a cozinha de hoje*. São Paulo, Editora Senac São Paulo.

GALE, W.A. e K.W. CHURCH (1991). A program for aligning sentences in bilingual corpora. In: *Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics*. Berkeley, California, p. 177-184.

GAUDIN, F. 1993. *Pour une socioterminologie: Des problèmes pratiques aux pratiques institutionnelles*. Rouen: Publications de l'Université de Rouen.

GELLERSTAM, M. (1986). Translationese in Swedish novels translated from English. In: WOLLIN, L. e H. Lindquist. (1986). *Translation Studies in Scandinavia*. Lund: CWK Gleerup, p. 88-95.

GEERAERTS, D. (1989). Prospects and Problems of Prototype Theory. *Linguistics*, vol. 27, p. 587-612.

_____ (1995). Representational formats in Cognitive Semantics. *Folia Linguistica* 39, p. 21-41.

GENTZLER, E. (1993). *Contemporary Translation Theories*. London and New York: Routledge.

GERBER, R. M. e V. VASILÉVSKI (orgs.) (2007). *Um percurso para pesquisas com base em corpus*. Florianópolis: Editora da UFSC.

GERLOFF, P.A. (1988). *From French to English: a look at the translation process in students, bilinguals, and professional translator*. PhD. Thesis Harvard University, Boston.

GINEZI, L.C. (2008). *Cafés do Brasil: estudo de variantes em português e inglês na língua falada*. Dissertação (Mestrado em Estudos Lingüísticos e Literários em Inglês) – Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo. Disponível em: <http://www.teses.usp.br/teses/disponiveis/8/8147/tde-03042008-134508/> [acesso em 10/10/08].

GOMENSORO, M. L. *Pequeno Dicionário de Gastronomia*. Rio de Janeiro: Editora Objetiva Ltda, 1999.

GOMES, L.G. e L. BARBOSA (2004). Culinária de papel. *Estudos Históricos* nº 33. Rio de Janeiro. Disponível em: <http://www.cpdoc.fgv.br/revista/arq/370.pdf> [acesso em: 24/04/08].

GÓMEZ González-Jover, A. (2005). *Terminografía, lenguajes profesionales y mediación interlingüística. Aplicación metodológica al léxico especializado de la industria del calzado y las*

industrias afines. Tese (Doutorado em Filologia Inglesa). Facultad de Filosofía y Letras da Universidade de Alicante, Alicante. Disponível em:

<http://www.cervantesvirtual.com/FichaObra.html?Ref=22151> [acesso em 10/10/08]

GÓMEZ González-Jover, A. e Ch. S. VARGAS (2004). Aspectos metodológicos para la elaboración de diccionarios especializados bilingües destinados al traductor. *El español, lengua de traducción. II Congreso Internacional*. Disponível na Internet no endereço:

<http://www.toledo2004.net/html/contribuciones/gomez-vargas.htm>, acesso em out. 2006.

_____. (2002). Córpora comparables y paralelos para la detección de terminología bilingüe: su explotación y uso con herramientas informáticas. *Actas del VIII Simposio de Riterm*. Disponível em: http://www.riterm.net/actes/8simposio/gomez_vargas.htm [acesso em: 05/5/08]

GONÇALVES, L.B. (2006). *'Dubliners' sob a lupa da lingüística de corpus: uma contribuição para a análise e a avaliação da tradução literária*. Tese (Doutorado em Estudos Lingüísticos e Literários em Inglês) – Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo. Disponível em: <http://www.teses.usp.br/teses/disponiveis/8/8147/tde-08112007-154609/> [acesso em 10/10/08]

GONSALVES, P.E. (2001). *Livro dos alimentos*. São Paulo: Summus.

GÖPPERICH, S. (1993). Die translatorische Behandlung von Textsortenkonventionen in technischen Texten. *Lebende Sprachen*. No. 2/93, pp.49-52.

GOUADEC, D. (1994). Nature et traitement des entités phraséologiques. *Terminologie et Phraséologie: Acteurs et Aménageurs. Actes du deuxième Université d'Automne en Terminologie*. Paris: La Maison du Dictionnaire, p. 164-193.

Gourmand World Cookbook Award (2005 edition). Disponível em:

http://www.cookbookfair.com/html/10_years_of_cookbook.html [acesso em: 19/8/2008]

GUTT, E-A. (1991). *Translation and relevance: Cognition and context*. Oxford: Blackwell.

HALLIDAY, M.A.K. (2001). Towards a theory of good translation. In: STEINER, E. e YALLOP, C. (eds.). *Exploring translation and multilingual text production: Beyond content*. Berlin and New York: Mouton de Gruyter, pp.13-18.

_____. (1993). Quantitative studies and probabilities in grammar. In: HOEY, M. (ed.). *Data Description Discourse -- Papers on the English Language in Honour of John McH Sinclair on his Sixtieth Birthday*. London: HarperCollins, p. 1-25.

_____. (1992). Language as system and language as instance: The corpus as a theoretical construct. In Jan Svartvik (Ed.), *Directions in Corpus Linguistics. Proceedings of Nobel Symposium 82*, Stockholm, 4-8 August 1991 (pp. 61-78). Berlin, New York: De Gruyter.

_____. (1991). Corpus studies and probabilistic grammar. In: AIJMER, K. e ALTENBERG, B. (eds.). *English Corpus Linguistics: Studies in Honour of Jan Svartvik*. London: Longman, p. 30-43.

HANKS, P. (2008). The Lexicographical Legacy of John Sinclair. *International Journal of Lexicography* 21(3), p. 219-229. Disponível em:

<http://ijl.oxfordjournals.org/cgi/content/abstract/21/3/219> [acesso em: 04/09/08]

HANSEN, S. (1988). Korpuslinguistik. Teori – metode – praksis. *LAMBDA no. 5*. Institut for Datalogivintik, Handelshøjskolen I København.

HATIM, B. e J. MUNDAY (2004). *Translation: An Advanced Resource Book*. London: Routledge.

HECK, M.C. (2004). Comer como atividade de lazer. *Estudos Históricos nº 33*. Rio de Janeiro, p. 1-22. Disponível em: <http://www.cpdoc.fgv.br/revista/arq/363.pdf> [acesso em: 20/04/08].

HELENE, H. (2006). *Dicionário de termos de gastronomia francês / português*. São Paulo: Gaia / Editora Bocatto.

HERBST, S. T. (2001). *The New Food Lover's Companion*. 3a. ed. New York: Barron's Educational Series, Inc.

HOEY, M. (2003). What's in a word? *English Teaching Professional, Issue 27*. Disponível em: <http://www.macmillandictionary.com/med-magazine/August2003/10-Feature-Whats-in-a-word.htm> [acesso em 07/7/04]

_____ (1997). From concordance to text structure: new uses for computer corpora. In: MELIA, J. e LEWANDOWSKA, B. (eds). *PALC '97: Proceedings of Practical Applications of Linguistic Corpora Conference*. Lodz: Lodz University Press, p. 2-23.

_____ (1993). *Patterns of lexis in text*. Oxford: Oxford University Press, 1991.

HOFFMANN, L. (2004 [1998]). Conceitos básicos da lingüística das linguagens especializadas. [tradução de: Ulla Pedde Muss e Maria José Borcomny Finatto]. In: KRIEGER e ARAÚJO (orgs.) (2004).

_____ (1991). Texts and text types in LSP. In: SCHRÖDER, H. (ed.). *Subject-oriented texts: languages for special purposes and text theory*. Berlin: de Gruyter.

_____ (1985). *Kommunikationsmittel Fachsprache*. Tübingen, G.Narr Verlag.

HOLMES, J. S. (1972) The Name and Nature of Translation Studies: expanded version. *Translated! Papers on Literary Translation and Translation Studies*. Amsterdam: Rodopi, 1988.

HOLZ-MÄNTTÄRI, J. (1984). *Translatorisches Handeln. Theorie und Methode*. Helsinki: Suomalainen Tiedeakatemia.

HÖNIG, H.G. e P. KUSSMAUL (1982). *Strategie der Übersetzung. Ein Lehr- und Arbeitsbuch*. Tübingen: Narr.

HUDSON, J. (1998). *Perspectives on fixedness*. Lund: Lund University Press.

HULLAH, E. (1984). *Cardinal's handbook of recipe development*. Toronto: Cardinal Biologicals Ltd.

HUMBLEY, J. (1997). Is terminology specialized lexicography? The experience of French-speaking countries. *Hermes, Journal of Linguistics no. 18*. Disponível em: http://hermes2.asb.dk/archive/FreeH/H18_02.pdf [acesso em 12/08/07]

- HUNG, E. e D. POLLARD (1998). Chinese Tradition. In: Baker, M. (ed.) (1998).
- HUNSTON, S. 2002. *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.
- HUNSTON, S. e G. FRANCIS (2000). *Pattern Grammar: A Corpus-driven Approach to the Lexical Grammar of English*. Amsterdã / Filadélfia: John Benjamins.
- JACOB, D. (2005). *Will write for food: The complete guide to writing restaurant reviews, cookbooks, recipes, stories, and more*. New York: Marlowe and Company.
- JAKOBSON, R. (1969 [1960]). Aspectos lingüísticos da tradução. *Lingüística e Comunicação*. São Paulo: Cultrix.
- JONES, B. *Dicionário Prático de Culinária*. [trad. Maria Emília de Oliveira]. São Paulo: Companhia Melhoramentos, 1996.
- JUNQUEIRA, L. (1980). *Ervas e especiarias na cozinha*. Rio de Janeiro: Tecnoprint / Ediouro.
- KAMP, D. (2006). *The United States of arrugula*. New York: Broadway Books.
- KENNY, D. (1998). Equivalence. In: BAKER (ed.) (1998), pp. 77-80.
- KINGSCOTT, G. (2002). Technical Translation and Related Disciplines. *Perspectives: Studies in Translatology*, Vol. 10, p. 247-255.
- KITTREDGE, R. (1982). Variation and homogeneity of sublanguages. In: KITTREDGE, R. e LEHRBERGER, J. *Sublanguage: Studies of language in restricted semantic domains*. Berlin / New York: Water de Gruyter.
- KJELLMER, G. (1992). Grammatical or nativelike? New Directions in English Language Corpora. In: LEITNER, G. *New Directions in English Language Corpora*. Berlin, Mouton de Gruyter, p. 329-344.
- KLIE, V. (2006). *Glossário de gastronomia: português-inglês / inglês-português*. São Paulo: Disal, série Easy Way.
- _____ (2003). *Not Just Hamburgers! Aprenda inglês com as melhores receitas da cozinha americana*. São Paulo: Disal, edição bilíngüe.
- KOLLER, W. (1995). The Concept of Equivalence and the Object of Translation Studies. *Target* 7(2), p. 191-222.
- KÖNIGS, F.G. (1992). Übersetzen im Fremdsprachenunterricht: Theoretische Erwägungen und praktische Anregungen. In: JUNG, U.H. (ed.). *Praktische Handreichungen für den Fremdsprachenlehrer*. Frankfurt: Lang.
- KORNING ZETHSEN, K. (1999). The dogmas of technical translation: Are they still valid? *Hermes, Journal of Linguistics* no. 23, pp. 65-75. Disponível em: http://hermes2.asb.dk/archive/FreeH/H23_05.pdf [acesso em 01/05/08]
- _____ (1997). *Expressivity in technical texts – from a translation theoretical perspective*. PhD dissertation. Århus: Handelshøjskolen i Århus, Det Erhvervssproglige Fakultet, Engelsk Institut.

KÖVESI, B., C. SIFFER, C. CREMA, G. MARTINOLI (2007). *400g: Técnicas de Cozinha*. São Paulo; Companhia Editora Nacional.

KRIEGER, M.G. (2006). Terminologia técnico-científica: políticas lingüísticas e Mercosul. *Revista Ciência e Cultura* vol.58, no.2 (abril/junho). São Paulo, p. 48-51. Disponível em: http://cienciaecultura.bvs.br/scielo.php?pid=s0009-67252006000200017&script=sci_arttext [acesso em: 30/07/06].

KRIEGER, M.G. e M.J.B. FINATTO (2004). *Introdução à Terminologia: Teoria & prática*. São Paulo: Contexto, 2004.

KRIEGER, M.G. e L. ARAÚJO (orgs.) (2004). *A terminologia em foco. Cadernos de Tradução* 17. Porto Alegre, Instituto de Letras da UFRGS, out.-dez.

KRIEGER, M.G. e A.M.B. MACIEL (orgs.) (2001). *Temas de Terminologia*. São Paulo e Porto Alegre: Ed. Universidade / Humanitas.

KRISHNAMURTHY, R. (1998). The Indian Tradition. In: BAKER, M. (ed.) (1998).

LAKOFF, G. (1987). *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*. Chicago: University of Chicago Press.

LAMPARELLI, A.H.C.A. (2007). *A naturalidade na tradução: quem garante?*. Dissertação (Mestrado em Estudos Lingüísticos e Literários em Inglês) – Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo. Disponível em: <http://www.teses.usp.br/teses/disponiveis/8/8147/tde-04122007-102315/> [acesso em 10/10/08]

LARA, L.F. (1999). Término y cultura: hacia una teoría del término. In: CABRÉ, M. T. (ed.) *Terminología y modelos culturales*. Barcelona: Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra, p. 39-60.

LAURÉN, C. (1998). Shift of perspective in scientific texts: A study of idiolect. In: LUNDQUIST et al. (eds.). *LSP Identity and interface. Research, knowledge and society. Proceedings of the 11th European Symposium on Language for Special Purposes*. p. 463-472.

LAURÉN, C. e PICHT, H. (orgs.). (1993). *Ausgewählte Texte zur Terminologie*. Wien: TermNet.

LAVIOSA, S. (2002). *Corpus-Based Translation Studies*. Amsterdam / New York: Rodopi.

LEECH, G. (1992). Corpora and theories of linguistic performance. In: SVARTVIK, J. (ed.). *Directions in corpus linguistics*. Proceedings of Nobel Symposium 82. Berlin: Mouton de Gruyter, pp. 8-29.

_____. (1991). The state of the art in Corpus Linguistics. In: AIJMER, K. e ALTENBERG, B. (eds.). *English Corpus Linguistics: Studies in Honour of Jan Svartvik*. London: Longman, p. 30-43.

LEFEVERE, A. (1990). Translation: its genealogy in the West. In BASSNETT, S. e A. Lefevere. *Translation, History and Culture*. New York: Cassell, p. 14-28.

L'HOMME, M.C. (2006). A Look at some Canadian Contributions to Terminology. In: Picht, H. (ed.). *Modern Approaches to Terminological Theories and Applications, Linguistic Insights*. Bern: Peter Lang, pp. 55-75.

_____. (2004). A Lexico-semantic Approach to the Structuring of Terminology. *Computerm 2004*. Disponível em: http://www.ling.umontreal.ca/lhomme/docs/Computerm_2004.pdf [acesso em 30/4/08].

LOPES, M.C. (2000). *Homepages institucionais em português e suas versões em inglês: um estudo baseado em corpus sobre aspectos lexicais e discursivos*. Dissertação (Mestrado em Lingüística Aplicada e Estudos da Linguagem) – Pontifícia Universidade Católica de São Paulo, São Paulo. Disponível em: http://www.pucsp.br/pos/lael/lael-inf/def_teses.html [acesso em: 10/10/08]

LÖRSCHER, W. (1991). *Translation performance, translation process, and translation strategies: A psycholinguistic investigation*. Tübingen: G. Narr.

MALMKJAER, K. (2005). *Linguistics and the language of translation*. Edinburgh: Edinburgh University Press.

_____. (1998). Unit of translation. In: BAKER, M. (ed) (1998).

MAIA, B. (2002). Do-it-yourself, disposable, specialised mini corpora – where next? Reflections on teaching translation and terminology through corpora. In: TAGNIN (org.) (2002c).

_____. (1997). Do-it-yourself corpora... with a little bit of help from your friends!. In: Barbara Lewandowska-Tomaszczyk, B. e Melia, P.J. (eds). *PALC '97 Practical Applications in Language Corpora*. Lodz: Lodz University Press, p. 403-410.

MARINAI, E. ; C. PETERS e E. PICCHI. (1991). Bilingual reference corpora: A system for parallel text retrieval. *Using corpora. Proceedings of the Seventh Annual Conference of the UW Centre for the New OED and Text Research* (Oxford, 29 Sept-1 Oct 1991), p 40-62.

MARTINS, J.C. (2007). *A terminologia do ecoturismo como espelho de diferentes visões: agências de turismo, ambientalistas e governo*. Dissertação (Mestrado em Estudos Lingüísticos e Literários em Inglês) – Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo. Disponível em: <http://www.teses.usp.br/teses/disponiveis/8/8147/tde-20022008-103947/> [acesso em: 10/10/08]

MATTAR, A. e P. REIS (2007). *Guia de boas compras de gastronomia 2007*. Rio de Janeiro: Editora Gryphus.

MEIJS, W. (1996). Linguistic Corpora and Lexicography. *Annual Review of Applied Linguistics*, vol. 16, p. 99-114.

MEL'ČUK, I., A. CLAS e A. POLGUÈRE (1995). *Introduction à la lexicologie explicative et combinatoire*, Coll. Champs linguistiques/Universités francophones, Louvain-la-Neuve/Paris, Éditions Duculot/AUPELF-UREF.

MELO, J. (1999). *Berinjela se escreve com j*. São Paulo: DBA.

- MONTORO, T. (2004). A construção de imagens da gastronomia nos cadernos de turismo dos jornais brasileiros. In: ARAÚJO *et al.* (org) (2004).
- MOON, R. (1998). *Fixed expressions and idioms in English: A corpus-based approach*. Oxford: Clarendon Press.
- MORAES, H.F.R. (2005). *O tradutor pode estar redondamente enganado: um estudo contrastivo de colocações adverbiais (inglês-português) sob o enfoque da lingüística de corpus*. Dissertação (Mestrado em Estudos Lingüísticos e Literários em Inglês) – Faculdade de Filosofia, Letras e Ciências Humanas da Universidade de São Paulo, São Paulo.
- MOURA, R. A. *Dicionário de Culinária e termos afins*. São Paulo: Atheneu, 2000.
- MOUNIN, G. (1963). *Les Problèmes théoriques de la traduction*. Paris, Gallimard.
- MYERS, G. (1989). The pragmatics of politeness in scientific articles. *Applied Linguistics* 10 (1). p. 1-35.
- NATTINGER, J.R. e J.S. DeCARRICO (1992). *Lexical phrases and language teaching*. Oxford: Oxford University Press.
- NEGRAES, M. (2003). *Guia A-Z de plantas: Condimentos*. São Paulo: Bei Comunicação.
- NERI, M. (2006). Patrimônio à mesa. *Desafios do desenvolvimento*, Ed. 26 (Dezembro de 2006). Brasília: IPEA. Disponível em: <http://www.ipea.gov.br/desafios/edicoes/26/artigo29609-4.php> [acesso em: 19/08/08].
- NEWMARK, P. (1988). *A Textbook of Translation*. Hemel Hempstead, UK: Prentice Hall.
- _____ (1969). Some notes on translation and translations. In: *The Incorporated. Linguist* 8 (4). p. 79-85.
- NORD, C. 1997. *Translating as a purposeful activity: functionalist approaches explained*. Manchester, UK: St. Jerome, 1997.
- _____ (1991). *Text analysis in translation*. Amsterdam: Rodopi.
- OLOHAN, M. (2007). The Status of Scientific Translation in Translation Studies. *Journal of Translation Studies*, 10(1): 131-144.
- _____ (2004). *Introducing corpora in translation studies*. London - New York: Routledge.
- ORENHA, A. (2004). *A compilação de um glossário bilíngüe de colocações, na área de Negócios, baseada em corpus comparável*. Dissertação (Mestrado em Estudos Lingüísticos e Literários em Inglês) – Faculdade de Filosofia, Letras e Ciências Humanas da Universidade de São Paulo, São Paulo.
- OSTMANN, B.G. e J.L. BAKER (2001). *The recipe writer's handbook*. New York: John Wiley & Sons, Inc.
- PARDO, R.M. (2004). *Critérios de construção e organização de um corpus de especialidade: o corpus técnico-científico de ortodontia*. Dissertação (Mestrado em Estudos Lingüísticos e

Literários em Inglês) – Faculdade de Filosofia, Letras e Ciências Humanas da Universidade de São Paulo, São Paulo.

PARTINGTON, A. (1998). *Patterns and meanings- using corpora for English language research and teaching*. Amsterdã: John Benjamins.

PAVEL, S. (2003). A Fraseologia na Língua de Especialidade: Metodologia de registro nos vocabulários terminológicos. In: FAUSTICH, E.; ABREU, S.P. (orgs.). *Lingüística Aplicada à Terminologia e à Lexicografia*. [tradução: Germano H. P. de Souza]. Porto Alegre: UFRGS, IL/UNB, p. 99-132.

PAWLEY, A. and F. SYDER (1983). Two puzzles for linguistic theory: nativelike selection and nativelike fluency. In: RICHARDS, J. and R. SCHMIDT (eds.). *Language and Communication*. London: Longman.

PEARSON, J. (2004 [1999]) Como ter acesso a elementos denfinitórios nos textos especializados?. [tradução: Carolina Huang e Sandra Dias Loguercio]. In: KRIEGER e ARAÚJO (orgs) (2004).

_____ (1998). *Terms in Context*. Amsterdam: John Benjamins.

PÉREZ Hernández, M.C. (2002). Explotación de los córpore textuales informatizados para la creación de bases de datos terminológicas basadas en el conocimiento. *Estudios de Lingüística del Español* 18. Disponível em: <http://elies.rediris.es/elies18/> [acesso em: 08/04/08].

PERROTI-GARCIA, A.J. e R.R. REBECHI (2007). *Vocabulário para Química*. Série Mil e Um Termos. São Paulo: SBS.

PETERSON, J.B. e D.C. PETERSON (1995). *Eat smart in Brazil*. Madison: Ginko Press.

PICHT, H. (2004 [2001]). Corpora como ponto de partida para a extração de dados termiológicos. [tradução: Danatela Duarte e Maria José Bocorny Finatto]. In: KRIEGER e ARAÚJO (orgs) (2004).

PINCHUCK, I. (1997). *Scientific and Technical Translation*. London: André Deutsch.

POSSAMAI, V. (2006). Marcadores Textuais de Textos Especializados em Tradução. *Tradterm*, v. 12, p. 159-179.

PUURTINEN, T. (1989). Two Translations in Comparison: A Study on Readability. In TIRKKONEN-CONDIT, S. e S. CONDIT. (eds.). *Empirical Studies in Translation and Linguistics*. Joensuu: University of Joensuu, 1989.

PYM, A. (1992). The Relations Between Translation and Material Text Transfer. *Target* 4:2.

QUIRK, R. (1960). Towards a description of English usage. *Transactions of the Philological Society*, pp. 40-61.

QUIRK, R., S. GREENBAUM, G. LEECH e J. SVARTVIK (1985). *A comprehensive grammar of contemporary English*. London: Longman.

- RAJAGOPALAN, K. (2007). Panorama da Lingüística de Corpus (LC). In: GERBER e VASILÉVSKI (2007).
- REISS, K. (1976). *Texttyp und Übersetzungsmethode*. Kronberg: Scriptor.
- REISS, K. e H. VERMEER (1984). *Grundlegung einer allgemeinen Translationstheorie*. Tübingen: Niemeyer.
- REVEL, J. *Um banquete de palavras: uma história da sensibilidade gastronômica*. [trad. Paulo Neves]. São Paulo: Companhia das Letras, 1996.
- RICARDO, L. (2007). Estante culinária. *Diário do Nordeste*, Fortaleza, 26/10/2007. Caderno 3, Comer & Beber. disponível em: <http://diariodonordeste.globo.com/materia.asp?Codigo=481951> [acesso em: 20/04/08]
- ROBINSON, D. (1997). *Western Translation Theory, from Herodotus to Nietzsche*. Manchester: St Jerome.
- ROCHA, M. (2007). Métodos estatísticos comuns em Lingüística de Corpus: visão geral. In: GERBER e VASILÉVSKI (2007).
- ROSCH, E.H. (1973). Natural Categories. *Cognitive Psychology* 4, p. 328-350.
- ROSSINI, A.M.Z.P. (2005). *A linguagem dos contratos bancários internacionais em inglês: um estudo descritivo baseado em Lingüística de Corpus*. Dissertação (Mestrado em Lingüística Aplicada e Estudos da Linguagem) – Pontifícia Universidade Católica de São Paulo, São Paulo. Disponível em: http://www.pucsp.br/pos/lael/lael-inf/teses/Rossini_2005_linguagem_contratos_bancarios.pdf [acesso em 10/10/08]
- SALDANHA, R.M. (2007). *Dicionário tradutor de gastronomia em seis línguas*. São Paulo: Antonio Bellini Editora & Cultura.
- SALGADO, A.R. (2006). *Unidades Fraseológicas Especializadas na perspectiva da tradução*. Dissertação. (Mestrado em Estudos da Linguagem) – Instituto de Letras, Universidade Federal do Rio Grande do Sul, Porto Alegre.
- SAGER, J.C. (1990). *A Practical Course in Terminology Processing*. Amsterdam and Philadelphia: John Benjamins.
- SAGON, C. (2006). Cooking 101: Add 1 Cup of Simplicity. *Washington Post*, 18/03/2006, page A01. Disponível em: <http://www.washingtonpost.com/wp-dyn/content/article/2006/03/17/AR2006031701969.html> [acesso em: 25/09/07].
- SÁNCHEZ, A. e P. CANTOS (1997). Predictability of word forms (types) and lemmas in linguistic corpora. A case study based on the analysis of the CUMBRE corpus: an 8-million word corpus of contemporary Spanish. *International Journal of Corpus Linguistics* Vol. 2(2), pp. 259-280.
- _____ (1996). *CUMBRE – Curso de Español*. Madri: SGEL.
- SCHMITT, P.A. (1999). *Translation und Technik*. Tübingen: Stauffenburg Verlag.

SCOTT, M. (1997). *WordSmith Tools*, version 3. Oxford: Oxford University Press.

SEBESS, M. (2008a). *Técnicas de cozinha profissional*. Rio de Janeiro: Senac Nacional.

_____ (2008b). *Técnicas de confeitaria profissional*. Rio de Janeiro: Senac Nacional.

SERRALVO BRIT, G.A. (2004). *Processo de versão de um artigo acadêmico de marketing com auxílio de corpora eletrônicos e da web*. Dissertação (Mestrado em Linguística Aplicada e Estudos da Linguagem) – Pontifícia Universidade Católica de São Paulo, São Paulo. Disponível em: http://www.pucsp.br/pos/lael/lael-inf/def_teses.html [acesso em: 10/10/08]

SINCLAIR, J. M. (2005). Corpus and text: basic principles. In: Wynne, Martin (ed.) *Developing Linguistic Corpora: A Guide to Good Practice*. Disponível em: <http://ahds.ac.uk/creating/guides/linguistic-corpora/chapter1.htm> [acesso em: 10/10/08]

_____ (1996). Preliminary recommendations on corpus typology. *Eagles Guidelines. Expert Advisory Group on Language Engineering Standards*. Disponível em: <http://www.ilc.cnr.it/EAGLES96/corpusyp/corpusyp.html> [acesso em: 10/10/08]

_____ (1995). From theory to practice. In: LEECH, G.; MYERS, G. e J. THOMAS (eds). *Spoken English on Computer*. Harlow: Longman, p. 99-112.

_____ (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

_____ (ed.). (1987). *Collins Cobuild English Dictionary*. London: Williams Collins Sons & Co.

SNELL-HORNBY, M. (1988). *Translation Studies: an Integrated Approach*. Amsterdam: John Benjamins.

STOLZE, R. (1982). *Grundlagen der Textübersetzung*. 2nd ed. Heidelberg: Julius Groos Verlag.

STUBBS, M. (2003). Two quantitative methods of studying phraseology in English. *International Journal of Corpus Linguistics* 7, 2.

_____ (2001). *Words and Phrases: Corpus studies of lexical semantics*. Oxford: Blackwell.

_____ (1996). *Text and corpus analysis. Computer-assisted studies of language and culture*. Oxford: Blackwell Publishers Ltd.

_____ (1995). Corpus evidence for norms of lexical collocation. In: COOK, G. e B. SEIDLHOFER (eds). *Principles and Practice in Applied Linguistics*. London: Oxford University Press.

TAGNIN, S. E. O. (2008). Disponibilização de corpora online: os avanços do Projeto COMET. In: TAGNIN, S.E.O. e O.A. VALE (orgs). *Avanços da Lingüística de Corpus no Brasil*. São Paulo: Humanitas, 2008, p. 101-121.

_____ (2007a). Lingüística de Corpus Aplicada a áreas afins: Projeto COMET. In: GERBER, e VASILEVSKI (2007).

_____ (2007b). A identificação de equivalentes tradutórios em corpora comparáveis. In: *Anais do I Congresso Internacional da ABRAPUI*: Belo Horizonte, 3 a 6 de junho de 2007.

_____. (2005). *O jeito que a gente diz: Expressões convencionais e idiomáticas*. São Paulo: Disal.

_____. (2002a). Os Corpora: Instrumentos de auto-ajuda para o tradutor. In: TAGNIN (org.) (2002c).

_____. (2002b). Corpora and the Innocent Translator: How can they help him. *inTRAlinea Special Issue: CULT2K*. Disponível em: http://www.intraline.it/specials/cult2k/eng_more.php?id=128_0_42_0_M% [acesso em: 27/4/08]

_____. (org.) (2002c). *Cadernos de Tradução no. 9, 2002/1*. Florianópolis: Núcleo de Tradução da Universidade Federal de Santa Catarina.

TAGNIN, S.E.O. e E.D. TEIXEIRA (2004a). British vs. American English, Brazilian vs. European Portuguese – how close or how far apart? A corpus-driven study. In: LEWANDOWSKA-TOMASZCZYK, B. (ed.). *Lodz Studies in Language vol. 9: Practical Applications in Language and Computers - PALC 2003*. Frankfurt am Main: Peter Lang. Disponível em: <http://www.fflch.usp.br/dlm/comet/> [acesso em: 01/10/2007]

_____. (2004b). Linguística de Corpus e Tradução Técnica - relato da montagem de um corpus multivarietal de culinária. *Tradterm*, São Paulo, v. 10, n. 1, p. 313-358, 2004.

TEIXEIRA, E.D. (2007). Etiquetagem em Linguística de Corpus: possibilidades de aplicações. In: GERBER e VASILEVSKI (2007).

_____. (2005). Corpus de Culinária. In: *CorTec – Corpus Técnico do Projeto COMET*. Disponível em: http://www.fflch.usp.br/dlm/comet/consulta_cortec.html [acesso em: 08/10/08]

_____. (2004). *Receita qualquer um traduz. Será? – a Culinária como área técnica de tradução*. Dissertação (Mestrado em Estudos Lingüísticos e Literários em Inglês) – Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2v.

_____. (2003). Em busca de um novo modelo tecno-formal para a construção de dicionários técnicos bilíngües – o exemplo da culinária. *Intercâmbio*, vol. XII, p. 243-251.

TEIXEIRA e TAGNIN (no prelo). *Vocabulário para Culinária inglês/português*. Série Mil e Um Termos. São Paulo: SBS.

TELINE, M.F. (2004) *Avaliação de métodos para a extração automática de terminologia de textos em português*. Dissertação (Mestrado em Ciências de Computação e Matemática Computacional) – Instituto de Ciências Matemáticas e de Computação da Universidade de São Paulo, São Carlos.

TEMMERMAN, R. (no prelo). Sociocultural situatedness of terminology in the life sciences: The history of splicing. In: FRANK, R. *et al. Body, Language and Mind. Vol II. Interrelations between Biology, Linguistics and Culture*. Tübingen: Springer Verlag.

_____. (2004 [2001]). Teoria Sociocognitiva da Terminologia [tradução: Natacha Enzweiler e Luiza Araújo]. In: KRIEGER e ARAÚJO (orgs) (2004).

_____ (2003). *Innovative methods in specialised lexicography*. Bergen: EBSCO Publishing. Disponível em: <http://www.hf.uib.no/forskerskole/review%20dancette.pdf> [acesso em 08/10/08].

_____ (2000). *Towards New Ways of Terminology Descriptio: The Sociocognitive approach*. Amsterdam: John Benjamins.

TEMMERMAN, R.; KERREMANS, K. e VANDERVOORT, V. (2005). La termontographie en contexte(s). In: BLAMPAIN, D.; THOIRON, P. e M. Van CAMPENHOUDT (eds). *Mots, Termes et Contextes. Actes des septièmes Journées scientifiques du réseau de chercheurs Lexicologie Terminologie Traduction*. Bruxelles, p. 429-439. Disponível em: <http://cvc.ehb.be/> [acesso em: 04/04/08].

TEMMERMAN, R. e KNOPS, U. (2004). The Translation of Domain Specific Languages and Multilingual Terminology Management. *Linguistica Antverpiensia New Series* 3. Antuérpia: Hogeschool Antwerpen.

TEUBERT, W. (2001). Corpus Linguistics and Lexicography. In: *International Journal of Corpus Linguistics*, vol. 6 - Special Issue, p. 125-153.

TOGNINI-BONELLI, E. (2001). *Corpus linguistics at work*. Amsterdam: John Benjamins.

_____ (2002). Functionally complete units of meaning across English and Italian: Towards a corpus-driven approach. In: ALTENBERG, B. e S. GRANGER (eds.). *Lexis in Contrast: Corpus-based approaches*. Amsterdam: John Benjamins, pp. 73–95.

TOURY, G. (1995). *Descriptive Translation Studies and Beyond*. Amsterdam e Philadelphia: John Benjamins.

_____ (1986). "Monitoring Discourse Transfer: A Test-Case for a Developmental Model of Translation". In House & Blum-Kulka (eds.). *Interlingual and Intercultural Communication. Discourse and Cognition in Translation and Second Language Acquisition Studies*. Tübingen: Gunter Narr, p.79-94

_____ (1980). *In Search of a Theory of Translation*. Tel Aviv: Porter Institute.

ULRYCH, M. (1997). The impact of multilingual concordancing on translation. In: LEWANDOWSKA-TOMASZCZYK, B. e J. MELIA (eds). *PALC '97. Practical Application in Language and Computers - PALC 1997*. Frankfurt e Main: Peter Lang.

UTZERI, F. (2003). O cardápio indigesto do tradutor. *JB online*. Disponível em http://jbonline.terra.com.br/index_frame.asp?jb/papel/colunas/fritz/2003/09/06/jorcolfri20030906001.html [acesso em: 07/09/03].

VALENTE, R.S. (2002). *La "Lexicologie explicative et combinatoire" dans le traitement des unités lexicales spécialisées*. Tese (doutorado em Lingüística) – Département de linguistique et de traduction, Faculté des arts et des sciences, Université de Montréal, Montréal. disponível em: http://www.theses.umontreal.ca/theses/nouv/valente_r-s/these.pdf [acesso em: 13/08/08]

VAN HOOFF, H. (1998). Os tradutores e os dicionários. In: DESLILE, J. e J. WOODSWORTH. *Os tradutores na história*. São Paulo, Ática.

VÁRADI, T. e G. KISS. (2001). Equivalence and Non-equivalence in Parallel Corpora. In: *International Journal of Corpus Linguistics* vol. 6 - Special Issue, p. 167-177.

VASILÉVSKI, V. (2007). Aspectos histórico-teóricos da Lingüística de Corpus: surgimento, abandono, levante e uso. In: GERBER e VASILÉVSKI (2007).

VARANTOLA, K. (2002). Disposable corpora as intelligent tools in translation. In: TAGNIN (Org.) (2002c).

_____. (1998). Translators and their Use of Dictionaries. User needs and user habits. In: ATKINS, B.T.S. (ed.). (1998), p. 179-192.

VENUTI, L. (1995). *The Translator's Invisibility*. London and New York: Routledge.

VERMEER, H. J. (1987). Literarische Übersetzung als Versuch interkultureller Kommunikation. In: WIERLACHER, A. (org.). *Perspektiven und Verfahren der interkulturellen Germanistik. Akten des I. Kongresses der Gesellschaft für Interkulturelle Germanistik*. München: Iudicium-Verlag, 541-549.

VIEIRA, E. V. e I. CÂNDIDO (2000). *Glossário Técnico Gastronômico, Hoteleiro e Turístico*. Caxias do Sul: EDUCS.

VINAY, J.P. e J. DARBELNET (1995 [1957]). *Comparative Stylistics on French and English: A methodology for translation*. Amsterdam / Philadelphia: John Benjamins.

VINTAR, Š. (2001). Using Parallel Corpora for Translation-Oriented Term Extraction. *Babel Journal*, John Benjamins Publishing.

_____. (2000). Extracting terminological collocations from a parallel corpus. *Fifth EAMT Workshop "Harvesting existing resources", May 11 - 12, 2000, Ljubljana, Slovenia*; pp.21-30. Disponível em: <http://www.mt-archive.info/EAMT-2000-Vintar.pdf> [acesso em: 08/10/08].

WHITMAN, J. e D. SIMON. (1993). *Recipes into type*. New York: HarperCollins.

WILSS, W. (1997). Hedging in expert-language reviews. In: Markkanen, R. e H. Schröder (eds.). *Hedging and Discourse: Approaches to the Analysis of a Pragmatic Phenomenon in Academic Texts*. Berlin and New York: de Gruyter.

_____. (1996). *Knowledge and skills in translator behaviour*. Amsterdam and Philadelphia: John Benjamins.

_____. (1982). *The Science of Translation: Problems and Methods*. Tübingen: Narr.

WITTGENSTEIN, L. (1987 [1953]). *Investigações Filosóficas*. Lisboa: Fundação Calouste Gulbenkian.

WRIGHT, S. E. (1993). The Inappropriateness of the Merely Correct: Stylistic Considerations in Scientific and Technical Translation. In: Sue Ellen Wright and Leland D. Wright Jr. (eds) *Scientific and Technical Translation*. Amsterdam: John Benjamins.

ZANATTA, R. (2004). *Tradução técnica e fraseologia: por um estudo integrado*. Dissertação (Mestrado em Língua e Literatura Alemã) - Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo.

ZIBART, E. (2001). *The Ethnic Food Lover's Companion: A Sourcebook for Understanding the Cuisines of the World*. Birmingham, Alabama: Menasha Ridge Press.

This document was created with Win2PDF available at <http://www.daneprairie.com>.
The unregistered version of Win2PDF is for evaluation or non-commercial use only.

Anexo I

Dados do Corpus Comparável de Culinária

A. Corpora de Inglês

Cód.	No. de arquivos /pastas	no. de receitas	tokens (sem et.)	types	t/t ratio
espec_IO	receitas variadas (39 arq. com várias receitas cada)	cerca de 350	69.440	4.796	6.91
Palc_IOB	<ul style="list-style-type: none"> • aco: side dishes (10 receitas) • ent: appetizers (10 receitas) • sal: salads (10 receitas) • ppm: entrées – pasta (10 receitas) • ppa: entrées – poultry (10 receitas) • ppc: entrées – meat (10 receitas) • ppp: entrées – fish and shellfish (10 receitas) • sop: soups (10 receitas) • sbr: desserts (10 receitas) • pao: bread (10 receitas) • teb: pies, tarts and cakes (10 receitas) 	110	27.974	2.513	8,98
Palc_IOA	idem Palc_IOB	110	30.231	2.836	9.38
HelenIOB	13 arquivos com várias receitas cada, nas categorias: <ul style="list-style-type: none"> • Starters (starts, snacks and suppers); • soups (soups, stocks, soup accompaniments); • meat (beef and veal; poultry; lamb and mutton; pork, gammon, ham and bacon; game; offal; mixed meat / other meat dishes); • fish; • vegetarian; • vegetables and accompaniments; • salads, • savoury sauces; • deserts; • confectionery; • drinks; 	1.465	261.852	6.272	2.40

	<ul style="list-style-type: none"> • Baking (cakes; biscuits, bars and slices; breads; scones; pastry) • Preserves (chutneys; jams, jellies and marmalades preserves) 				
BBC_recs	1 arquivo com 5.180 receitas em categorias variadas	5.180	1.146.241	9.259	0.81
fishSoup	1 arquivo com receitas variadas de peixes e frutos do mar	181	42.387	3.227	7.61
Total		cerca de 7.400	1.578.125	28.903	0.93

A. Corpora de Português

Cód.	arquivos /pastas	no. de receitas	tokens	types	t/t ratio
CDCC	<ul style="list-style-type: none"> • lan: lanches (100 receitas) • ppm: massas & molhos (100 receitas) • ppp: peixes & frutos do mar (100 receitas) • ppc: carnes & aves (100 receitas) • sob: sobremesas (100 receitas) • teb: tortas & bolos (100 receitas) 	600	131.966	2.804	2,12
CD_Rec	<ul style="list-style-type: none"> • aco: acompanhamentos (40 receitas) • ave: aves (80 receitas) • bol: bolos (38 receitas) • car: carnes (64 receitas) • chu:churrasco (54 receitas) • drk: drinks (50 receitas) • fnd: fondues, musses e soufflés (56 receitas) • mss: massas (82 receitas) • pao: pães (55 receitas) • ptp: típicos (28 receitas) • pxs: peixes (49 receitas) • sal: saladas (92 receitas) • snd: sanduíches e salgados (148 receitas) • sob: sobremesas e doces (68 receitas) • sop: sopas (40 receitas) • srv: sorevetes (28 receitas) • tor: tortas (43 receitas) 	1.015	133.340	3.809	2,86

palc_POB	<ul style="list-style-type: none"> • aco: acompanhamentos (10 receitas) • ent: entradas (10 receitas) • sal: saladas (10 receitas) • ppm: pratos principais – massas (10 receitas) • ppa: pratos principais – aves (10 receitas) • ppc: pratos principais – carnes (10 receitas) • ppp: pratos principais – peixes e frutos do mar (10 receitas) • sop: sopas (10 receitas) • sbr: sobremesas (10 receitas) • pao: pães (10 receitas) • teb: tortas e bolos (10 receitas) 	110	24.816	2.066	8,33
Basilico	“Receitas com...” (430 receitas) “Menu Basilico” + “Receitas do dia” + “Famosos na cozinha” + “À mesa com Mari” (81 arquivos – em média 3 receitas por arq.)	cerca de 700	193.442	6.376	3,30
JPAN	receitas variadas de (1 arq)	243	41.164	3.355	8,15
GRGula	receitas variadas (1 arq.)	cerca de 45	14.046	1.626	11.58
Espec_PO	receitas variadas (23 arq., com várias receitas cada)	cerca de 600	102.833	3.907	3,80
Terra	receitas variadas – Acompanhamentos, Aves, Bebidas, Carnes, Frutos do mar, Grãos e massas, Lanches, Molhos, Saladas, Sobremesas: Bolos, Sobremesas: Tortas, Sobremesas: Doces, Sopas, Tortas, pizzas e pães (1 arq.)	5024	939.653	8.500	1.0
Total		cerca de 8.300	1.581.260	32.443	0,85

This document was created with Win2PDF available at <http://www.daneprairie.com>.
The unregistered version of Win2PDF is for evaluation or non-commercial use only.

Anexo II

Palavras-chave do subcorpus de inglês

	Culinária vs. ANC (written)	Culinária vs. Brown Corpus
1	TBSP	MINUTES
2	MINUTES	TBSP
3	MINS	ADD
4	ADD	MINS
5	OIL	OIL
6	UNTIL	UNTIL
7	CHOPPED	CHOPPED
8	PAN	PAN
9	OZ	OZ
10	BUTTER	BUTTER
11	INGREDIENTS	INGREDIENTS
12	HEAT	HEAT
13	SALT	SALT
14	COOKING	SUGAR
15	PEPPER	PEPPER
16	COOK	COOKING
17	SUGAR	COOK
18	TSP	TSP
19	SERVES	AND
20	SAUCE	METHOD
21	SERVE	SERVES
22	METHOD	SERVE
23	CREAM	CREAM
24	PREPARATION	SAUCE
25	OVEN	TIME
26	FLOUR	PREPARATION
27	GARLIC	OVEN
28	OLIVE	FLOUR
29	STIR	GARLIC
30	JUICE	OLIVE
31	ONION	STIR
32	MIXTURE	INTO
33	REMOVE	JUICE
34	BOWL	MIXTURE
35	LEMON	BOWL
36	FRESH	REMOVE
37	INTO	PLACE
38	EGG	FRESH
39	PLACE	EGG
40	TIME	LEMON
41	AND	ONION
42	POUR	MIX
43	CHICKEN	LESS
44	FINELY	WATER
45	FRESHLY	CHICKEN
46	WATER	FRESHLY
47	GROUND	POUR
48	SLICED	FINELY
49	MIX	GROUND
50	MILK	MILK

This document was created with Win2PDF available at <http://www.daneprairie.com>.
The unregistered version of Win2PDF is for evaluation or non-commercial use only.

Anexo III

Palavras-chave do subcorpus de português

	Culinária vs. Banco de Português	Culinária vs. Lácio-Ref
1	SOPA	SOPA
2	SAL	SAL
3	CHÁ	CHÁ
4	COLHERES	COLHERES
5	PREPARO	XÍCARA
6	XÍCARA	MINUTOS
7	INGREDIENTES	PREPARO
8	COLHER	COLHER
9	COLOQUE	INGREDIENTES
10	MANTEIGA	COLOQUE
11	MINUTOS	FOGO
12	MISTURE	MANTEIGA
13	RETIRE	LEITE
14	DEIXE	DEIXE
15	PANELA	ÁGUA
16	FOGO	DE
17	FORNO	FARINHA
18	FARINHA	LEVE
19	JUNTE	AÇÚCAR
20	AZEITE	MOLHO
21	ÁGUA	MISTURE
22	MOLHO	PANELA
23	LEITE	AZEITE
24	TEMPERATURA	RETIRE
25	CEBOLA	ATÉ
26	AÇÚCAR	JUNTE
27	G	FORNO
28	CREME	CEBOLA
29	LEVE	CREME
30	ACRESCENTE	G
31	MASSA	MASSA
32	DE	ACRESCENTE
33	PIMENTA-DO-REINO	ALHO
34	ALHO	TEMPERATURA
35	SIRVA	QUENTE
36	DIFICULDADE	DIFICULDADE
37	MODO	GOSTO
38	QUENTE	E
39	COZINHE	COM
40	RESERVE	SIRVA
41	BATA	PIMENTA-DO-REINO
42	PORÇÕES	CATEGORIA
43	LIMÃO	RESERVE
44	PICADA	BEM
45	ADICIONE	LIMÃO
46	TRIGO	TRIGO
47	E	MODO
48	CATEGORIA	PIMENTA
49	GOSTO	COZINHE
50	TIGELA	BATA

This document was created with Win2PDF available at <http://www.daneprairie.com>.
The unregistered version of Win2PDF is for evaluation or non-commercial use only.

Anexo IV

Equivalentes para os padrões de “flour*” em dicionários

termo - corpus	dic. bilíngües	comentário / outros dicionários	internet ¹	oc.
barley flour	barley – cevada (Moura, Klie)	“farinha de cevada” – não consta de nenhum dos dicionários de culinária. Embora haja várias ocorrências do termo na Internet, não ocorre em receitas, mas em sites institucionais (empresas produtoras de farinha, padarias, etc.) ou em artigos sobre nutrição e história da alimentação. É, portanto, produto de pouco uso na culinária caseira brasileira, muito embora esteja disponível para compra nos mercados brasileiros (marca Ferla, por exemplo)	“farinha de cevada” "farinha de cevada" sal OR açúcar "farinha de cevada" receita	172 89 12
cornflour	corn flour – farinha de milho corn starch – amido de milho (Moura) cornstarch – amido de milho; maisena (Klie)	conforme comentado anteriormente, o termo “cornflour” tem significados diferentes nas culturas americana e britânica – o que não é sequer mencionado pelo dicionário bilíngüe, embora conste da maioria dos dicionários monolíngües de inglês. “ corn flour – (...) In British recipes the term ‘cornflour’ is used synonymously with the U.S. word corstarch” (NFLC). Portanto, o equivalente correto é “amido de milho” (PDGU; GTGH, DACB, PDGA – somente ‘amido’; DPCM – ‘farinha de amido de milho’, no verbete FARINHAS), ou “maizena” / “maisena” (PDGU, DACB, GTGH, PDGA) – designação metonímica pela qual ficou conhecido o ingrediente, originalmente grafado com “z”. DACB aponta “biscornil”, outro termo metonímico, como sinônimo.	“amido de milho” "amido de milho" sal OR açúcar "amido de milho" receita maisena maisena sal OR açúcar site:.br maisena receita maizena maizena sal OR açúcar maizena receita biscornil	15.800 10.900 5.540 14.400 10.200 5.270 14.900 9.570 3.430 7

¹ Todas as pesquisas foram feitas no Google, no dia 18/01/2005, usando a restrição “site:.br”, para limitar as buscas a sites brasileiros.

DPCM – Dicionário Prático de Culinária Melhoramentos (Jones, 1996); **PDGU** – Pequeno Dicionário da Gula (Algranti, 2000); **PDGA** – Pequeno Dicionários de Gastronomia (Gomensoro, 1999); **GTGH** – Glossário Técnico Gastronômico, Hoteleiro e Turístico (Vieira e Cândido, 2000); **DACB** – Dicionário-Almanaque de Comers e Bebes (Fornari, 2001); **NFLC** – New Food Lover’s Companion (Herbst, 2001)

floury potatoes	floury - farinhento	o termo “farinhento” foi encontrado com entrada apenas no DACB. No entanto, uma consulta ao termo “batata” revelou o uso de “farinhosa” no DPCM, no subverbo de BATATAS, King Edward – “(...) A polpa é branca e farinhosa”.	batata farinhenta batata farinhosa	13 12
gram flour	gram – grão-de-bico	“farinha de grão-de-bico” é mencionada pelos dicionários PDGA, PDGU e DPCM, dos quais DPCM e PDGA apontam “besan” como sinônimo. Do dicionário GTGH consta apenas o termo “besan”.	“farinha de grão-de-bico” “farinha de grão-de-bico” sal OR açúcar *besan ² *besan “grão-de-bico” *besan açúcar OR sal	53 22 211 0 21
granary flour	NÃO CONSTA	de acordo com várias fontes consultadas na internet, “granary flour” é termo metonímico usado para designar um tipo de farinha muito usado na Irlanda: “Granary is a brand of brown flour with malted wheat grains. It is milled by Rank Hovis Limited, ‘the UK’s leading flour miller.’ (...) we don’t know of anything similar available in the United States. You might try experiment with mixing whole-wheat flour with a little all-purpose flour and a bit of malt powder in the hopes of approximating Granary flour on your own” ³ ; “Granary is a registered trade name of Rank Hovis and is a malted wheat grain flour. Malted wheat grain flour is a brown flour with malted wheat grains added for distinctive nutty flavour” ⁴ . No DPCM consta, no verbete FARINHAS, o subverbo “Farinha de trigo com malte – geralmente é a farinha escura misturada com malte,	“farinha de trigo com malte” “farinha de malte” “farinha com malte” “farinha maltada”	0 11 0 2

² Apesar do número relativamente alto de ocorrências para “besan” em sites brasileiros, uma consulta aos resultados revelou que se tratava, em sua maioria, de letras de música ou poemas em espanhol em que a palavra é a forma flexionada do verbo “besar” (beijar, em português). Foram feitas então as outras buscas, nas quais não se obteve resultados satisfatórios – “besan açúcar OR sal” retornou 21 ocorrências, mas nenhuma delas relacionada à Culinária.

³ <http://www.ochef.com/746.htm>

⁴ <http://www.flourandgrain.com/explorers/grainChainClub/faqs.htm>

DPCM – Dicionário Prático de Culinária Melhoramentos (Jones, 1996); **PDGU** – Pequeno Dicionário da Gula (Algranti, 2000); **PDGA** – Pequeno Dicionários de Gastronomia (Gomensoro, 1999); **GTGH** – Glossário Técnico Gastronômico, Hoteleiro e Turístico (Vieira e Cândido, 2000); **DACB** – Dicionário-Almanaque de Comer e Beber (Fornari, 2001); **NFLC** – New Food Lover's Companion (Herbst, 2001)

		o que lhe confere textura e sabor de nozes”. Também no verbete “farinha de trigo” do PDGA consta: “A branca é também encontrada acrescida de algum outro ingrediente, como fermento, glúten, malte (...)”. Apesar de haver algumas ocorrências de “farinha de malte” na Internet brasileira, são em sites da Anvisa e de panificadoras, o que prova que é um produto usado apenas na Culinária industrial		
malted brown flour	NÃO CONSTA	parassinônimo de “granary flour” – “ Malted Brown flour is a mixture of brown flour, and grains of malted wheatflakes” ⁵		
malthouse flour	NÃO CONSTA	parassinônimo de “granary flour”, com o acréscimo de centeio na composição – “Organic Malthouse Flour 1kg, A blend of brown flour with soft malted grains, malt and rye flours for really tasty granary style breads” ⁶ ; “Organic Specialty Malthouse Flour – 1kg: A malty bread flour with wheat flakes and rye flour” ⁷		
plain flour	plain flour – farinha de trigo comum (Moura)	o DPCM, na entrada FARINHAS, contém um verbete intitulado “Farinhas de trigo para diferentes usos”, em que consta, como subverbo, “ Comum – significa que não recebeu nenhum ingrediente especial. Alguns livros de receita não especificam o tipo de farinha a ser usado (cuidado: as farinhas diferem de acordo com o tipo do trigo, e os resultados podem ser bastante diferentes, principalmente em assados)”. GTGH, PDGA e PDGU citam apenas “farinha de trigo”, como termo genérico para todas as farinhas obtidas do trigo, e DACB só possui o verbete genérico “farinha”.	"farinha de trigo comum" 324 "farinha de trigo comum" açúcar OR sal 263 "farinha de trigo" -integral 79.700 "farinha de trigo" açúcar OR sal -integral 63.800	

⁵ www.mcdougalls.com/basics/flour.htm

⁶ www.dovesfarm.co.uk/products-dietcompatibility.htm

⁷ <http://www.goodnessdirect.co.uk/cgi-local/frameset/detail/431679.html>

plain wholemeal flour	whole meal bread – pão integral (Moura)	por essa entrada é possível deduzir que “wholemeal flour” seja farinha de trigo integral, resta saber a que “plain” se refere. No site da empresa Allinson ⁸ , encontramos a seguinte definição: “This flour contains a lower amount of protein and therefore produces dough that is far less elastic and ‘short’ in nature. It is used to give the distinctive texture of cakes, scones and pastries. It is not suitable for bread making”. Num documento .pdf da empresa Carrs ⁹ , a definição é: “100% stoneground, contains 100% of the wheatgrain and high in natural fibre. Excellent for crumbles, biscuits and pastry.” Encontramos também na Internet as expressões “farinha de trigo integral grossa” e “farinha de trigo integral fina”, mas não conseguimos nenhuma informação quanto às suas diferenças físicas e de uso em receitas.	“farinha de trigo integral” “farinha de trigo integral” sal OR açúcar “farinha integral” “farinha integral” sal OR açúcar “farinha de trigo integral fina” “farinha integral fina” “farinha de trigo integral grossa” “farinha integral grossa”	1.910 1.570 1.270 775 148 40 18 17
potato flour	potato flour – amido de batata (fécula) (Moura)	PDGA – só “amido” e “fécula – amido que pode ser obtido da <i>batata</i> (...)”; PDGU – consta somente o “amido de milho”, e “fécula de batata”; DTGH – constam apenas os termos genéricos “amido – (...) obtido de plantas e grãos de cereais como trigo, milho, arroz, <i>batata</i> (...)” e “fécula”; DACB – constam “amido – (...) encontra-se, também, nos feculentos, como a <i>batata</i> (...)” e o genérico “fécula”	"fécula de batata" "amido de batata"	1.110 146
rice flour	rice – arroz (Moura, Klie) flour – farinha (Moura, Klie)	na direção português-inglês do dicionário bilíngüe consta o verbete “farinha de arroz – ground rice”. Uma pesquisa na internet parece indicar que trata-se de uma variação cultural entre americanos e britânicos ¹⁰ : “In this country [United States], ‘rice flour’ is our one form of ground rice. (...) Rice flour is	“farinha de arroz” "farinha de arroz" sal OR açúcar "creme de arroz" "creme de arroz" sal OR açúcar	687 963 659 446

⁸ http://www.allinsonflour.co.uk/rootpops/p_flours_culin_whpf.cfm

⁹ www.fabflour.co.uk/pdf/pub/22072004161448.pdf

¹⁰ <http://www.ochef.com/771.htm>

		used in baking and also as a thickening agent. (...) Ground rice in the UK is only slightly more varied, appearing in two forms — rice flour and an even finer rice powder. Use rice flour for the ground rice specified in your British recipe”. Em sites brasileiros, há menção a vários tipos de farinha de arroz, especialmente em receitas da culinária vegetariana / macrobiótica e indiana. Nos dicionários consultados, encontramos “ farinha de arroz – use-a para fazer alguns tipos de macarrão à moda oriental e rosquinhas. Serve também para biscoitos”, como subverbo de FARINHA no DPCM. No GTGH, consta “farinha de arroz”, cuja definição remete a “fubá de arroz – o mesmo que farinha de arroz ou creme de arroz. Arroz branco muito finamente moído, usado para engrossar e fazer muitos tipos de doces”. No PDGU e no PDGA constam apenas “farinha de arroz”, sendo que nesse último, consta também “creme de arroz”, embora não haja referência cruzada.	"fubá de arroz" "fubá de arroz" sal OR açúcar	37 35
seasoned flour	seasoning – tempero (Moura) (to) season – temperar seasoned salt – sal com temperos (Klie)	conforme observamos nas pesquisas que realizamos na Internet, “farinha temperada” não é um produto que possa ser encontrado nas prateleiras dos supermercados, mas uma mistura de farinha com temperos (especialmente sal e pimenta-do-reino) usada para enfarinhar carnes/aves/peixes antes da fritura ou cozimento – uma técnica aparentemente muito mais usada na	"farinha temperada" "farinha temperada" sal OR açúcar	170 88

DPCM – Dicionário Prático de Culinária Melhoramentos (Jones, 1996); **PDGU** – Pequeno Dicionário da Gula (Algranti, 2000); **PDGA** – Pequeno Dicionários de Gastronomia (Gomensoro, 1999); **GTGH** – Glossário Técnico Gastronômico, Hoteleiro e Turístico (Vieira e Cândido, 2000); **DACB** – Dicionário-Almanaque de Comidas e Bebidas (Fornari, 2001); **NFLC** – New Food Lover's Companion (Herbst, 2001)

		culinária britânica / americana que na brasileira		
self raising flour self-raising flour	self raising flour – farinha que já tem fermento (Moura)	o equivalente fornecido pelo dicionário bilíngüe é mais uma definição do que uma equivalência. A “farinha que já tem fermento” surgiu há pouco tempo no mercado brasileiro e é chamada, pelos fabricantes ¹¹ , de “farinha de trigo especial com fermento” (Dona Benta, Pão de Açúcar, p. ex.) “farinha de trigo com fermento” (Rosa Branca, Boa sorte, p. ex.);	"farinha de trigo com fermento" "farinha de trigo especial com fermento" "farinha com fermento"	235 31 150
self raising wholemeal flour	NÃO CONSTA	versão da farinha integral acrescida de fermento, sem equivalente no Brasil. A proporção para substituição é “One cup of self-rising flour contains 1 1/2 teaspoons baking powder and 1/2 teaspoon salt” ¹² . Portanto, nas receitas que pedirem esse ingrediente deve constar, “para cada xícara de farinha de trigo integral, acrescente uma colher e ½ de sobremesa de fermento (7,5g) e 1 colher de café de sal (2,5g)”.	"farinha de trigo integral com fermento" "farinha integral com fermento"	1 0
soft grain flour	(to) soften – amolecer; deixar amolecer (Klie)	o termo “soft grain flour” teve baixa ocorrência na internet (menos de 20). Encontramos a seguinte definição na internet: “ Soft wheat flours – grown in moderate climates are low in protein and therefore low in gluten strength. Soft wheat flours are used for chemically leavened (baking powder and baking soda) recipes, such as cakes, cookies and biscuits. (...) are perfect for anything that needs to be tender.” ¹³ . Encontramos apenas uma ocorrência de “farinha de trigo brando” na internet, em um resumo de artigo científico. De acordo com Baccin (2003,	"farinha de trigo especial" "farinha de trigo especial" sal OR açúcar "farinha de trigo de primeira" ¹⁴ "farinha de trigo de primeira" sal OR açúcar	2.840 2.460 14 14

¹¹ pesquisa feita no site Buscapé, que compara preços de produtos em diferentes supermercados: www.buscapede.com.br/

¹² <http://www.baking911.com/disc10/00000c60.htm>

¹³ http://www.baking911.com/pantry_flour,grains2.htm

¹⁴ apesar de termos encontrado 14 ocorrências de “farinha de trigo de primeira”, observamos que, mesmo com o uso do restritor “sal OR açúcar”, os resultados exibidos pelo Google não eram receitas, mas descrição de produtos.

		v.2: 275), na Itália são produzidos cinco tipos diferentes de farinha a partir do “grano tenero” (trigo brando ou “soft”), mas no Brasil a única distinção é entre a farinha de trigo (comum) – que corresponde à “farina di grano tenero tipo 1” e “tipo2” – e a “farinha de trigo especial” ou “de primeira” – que corresponde à “farina di grano tenero tipo 0”. Ao que tudo indica, “soft grain flour” seria o equivalente para a forma italiana “farina di grano tenero tipo 00” (o verbete farinha “0 0” consta do PDGU, mas diz que “(...) é produzida na cidade de Bolonha, Itália (...)”), sem equivalente no português – o que mais se aproximaria seria a “farinha de trigo especial”		
sponge cake flour	sponge cake – bolo esponja, pão-de-ló (Moura)	também com poucas ocorrências na internet (25), “sponge cake flour” é, aparentemente, parassinônimo de “cake flour”, ou “soft grain flour” – como discutido no termo anterior, o equivalente mais próximo em português seria “farinha de trigo especial”		
strong plain flour	NÃO CONSTA	“strong plain flour”, de acordo com algumas definições encontradas na internet, é a farinha especial para pães ou “farinha de trigo (comum) – com alto teor de glúten: BREAD FLOUR: Bread flour, also referred to as “strong white flour” or “strong flour”, is a high-gluten flour usually milled from hard wheat. It contains a high percentage of protein which forms gluten when moistened” ¹⁵ ; “bread flour = hard-wheat flour = strong flour = high-gluten flour. Notes: This flour has a high level of gluten, which gives bread more structure.” ¹⁶		
strong white bread	NÃO CONSTA	conforme sugerido pela definição citada acima,		

¹⁵ http://www.baking911.com/pantry_flour.grains.htm

¹⁶ <http://www.foodsubs.com/Flour.html#bread>

flour		parassinônimo de “strong plain flour” e “strong white flour” – “farinha de trigo (comum)” em português		
strong white flour	NÃO CONSTA	conforme sugerido pela definição citada acima, parassinônimo de “strong plain flour” e “strong white bread flour” – “farinha de trigo (comum)” em português		
strong wholemeal flour	NÃO CONSTA	versão integral da farinha de trigo comum, com alto teor de glúten – “farinha de trigo integral”		
unbleached plain flour	NÃO CONSTA	segundo definição encontrada na internet, “ Unbleached flour is aged and bleached naturally by oxygen in the air. It is more golden in color, generally more expensive and may not have the consistency in baking qualities that bleached flour does. Unbleached is preferred for yeast breads because bleaching affects gluten strength”. Não encontramos na internet qualquer referência a esse tipo de farinha no Brasil. Por ser a farinha de trigo comum a que possui tonalidade mais amarelada, acreditamos que “farinha de trigo (comum)” seja o equivalente mais próximo em português.		
unbleached plain white flour	NÃO CONSTA	parassinônimo de “unbleached plain flour” e “unbleached white bread flour” – “farinha de trigo (comum)”		
unbleached white bread flour	NÃO CONSTA	parassinônimo de “unbleached plain flour” e “unbleached plain white flour” – “farinha de trigo (comum)”		
wheat flour	wheatflour – farinha de trigo (Moura) wheat bran; bread; germ – farelo; pão; gérmén de trigo (Klie)	termo genérico para designar as farinhas provenientes do trigo (por oposição às farinhas provenientes de outros grãos) – usado em descrições, especialmente na introdução / comentário das receitas – “farinha de trigo”		

DPCM – Dicionário Prático de Culinária Melhoramentos (Jones, 1996); **PDGU** – Pequeno Dicionário da Gula (Algranti, 2000); **PDGA** – Pequeno Dicionários de Gastronomia (Gomensoro, 1999); **GTGH** – Glossário Técnico Gastronômico, Hoteleiro e Turístico (Vieira e Cândido, 2000); **DACB** – Dicionário-Almanaque de Comidas e Bebidas (Fornari, 2001); **NFLC** – New Food Lover's Companion (Herbst, 2001)

wheatmeal flour	wheat – trigo (Moura) Klie – idem acima	parassinônimo de “wholemeal flour” e “wholewheat flour” – “farinha de trigo integral (grossa / fina)”		
wholemeal bread flour	whole meal bread – pão integral	parassinônimo de “wholemeal flour” e “wheatmeal flour” – “farinha de trigo integral (grossa / fina)”		
wholemeal flour	NÃO CONSTA	de acordo com definição encontrada na internet: “wholemeal flour – Also known as wholewheat or wheatmeal flour and is made from the whole wheat grain” ¹⁷ ; “Wholemeal, also called whole wheat flour contains 100% of the wheat grain including the bran and the germ with nothing added or taken away. It is rich in fiber to add roughage to the diet” ¹⁸ . Parassinônimo de “wholemeal bread flour”, “wholewheat flour” e “wheatmeal flour”.		
wholemeal self raising flour	NÃO CONSTA	parassinônimo de “self raising wholemeal flour”	farinha integral com fermento	3
wholewheat flour	whole wheat flour – farinha integral (Klie)	parassinônimo de “wholemeal flour” e “wheatmeal flour”		

¹⁷ <http://www.wheat-free.org/wheat-flour.html>

¹⁸ http://www.baking911.com/pantry_flour,grains2.htm

This document was created with Win2PDF available at <http://www.daneprairie.com>.
The unregistered version of Win2PDF is for evaluation or non-commercial use only.

Anexo V

Dados sobre a metodologia semi-automática de extração

padrão de busca	oc. tt no corpus	types ≥ 3 ocs	oc. dos types ≥ 3	exemplos de UTEs	exemplos de não-UTEs
adj + adv	878	51	487	ralado grosso moída grosseiramente picadas grosseiramente	fervente levemente fresco bem pequena bem
adj + adv + adj	243	18	82	cortado bem fininho picado bem fino	fresco bem gelado grande bem picada média bem picada
adj + adv + adv	16	0	0	-	médias não muito preta já bem ensopado ainda mais
adj + conj + adj	1356	97	672	lavado e escorrido lisa e homogênea quente ou fria fundo e laterais	cozido e cortado limpo e fervido pequenos e limpos
adj + prp + adv	365	28	246	-	brando por mais Cuidado para não preaquecido por cerca dissolvida num pouco
adj + prp + adv + n	8	0	0	-	média com bastante óleo individuais com bastante manteiga verde por mais tempo
adj + prp + art + n	1325	64	390	forrada com um pano rasgadas com as mãos picadas com as mãos	fundo de uma travessa untadas com o óleo baixo com a panela

adj + prp + n	8236	530	4.960	moída na hora untada com manteiga polvilhada com farinha cortados ao meio	fervente com sal médios sem pele químico em pó média em cubos
adj + prp + n + adj	1216	92	497	picada em pedaços pequenos cortado em cubos pequenos cortado em tiras finas cortadas em rodela grossas	média em cubos pequenos média em rodela finas redonda de aro removível médio de ervas frescas
adj + prp + num	1571	145	923	cortados em 4 cortados em quatro cortadas em quatro	médio por 4 manual por 10 morna em duas picadas de 2 pré-aquecido por 10
subtotal	15.214	1.025	8.257		
adv + adj	3760	232	3.009	ainda quente bem picadinha levemente dourada grosseiramente picadas	bem vermelhos mais dura bem miudinho pouco mole
adv + adv + adj	153	15	79	não muito grossas não muito maduros não muito pequenos ainda bem quente	também bastante versátil
adv + art + n + adj	166	11	47	só a parte branca somente a parte branca só a parte central	bem uma frigideira antiaderente assim uma massa homogênea bem os funghi secchi finamente a cebola restante
adv + art + n + v	292	16	102	enquanto o forno aquece	Quando a manteiga estiver Quando a carne estiver quando o grão ficar
adv + art + n + v + v	33	1	3	Quando a manteiga estiver borbulhando	-
adv + conj + adv	79	2	63	mais ou menos	sempre e não

adv + conj + v	1630	112	1.238	assim que ferver assim que aquecer assim que derreter assim que amornar	bem e leve como se fosse bem e acrescente bem e deixe
adv + n	1529	90	1.186	demais ingredientes como acompanhamento bastante gelo	menos gordura menos sal muito cuidado também ingrediente
adv + pron	528	16	447	Enquanto isso	mais alguns bem todos mais uns quase todas
adv + prp + num + n	1729	88	1.438	cerca de 30 minutos depois de 3 minutos cerca de 12 kg cerca de 2 colheres	levemente dos dois lados delicadamente por 1 minuto
adv + prp + v	961	62	641	antes de servir antes de assar só para gratinar somente para aquecer	bem para envolver delicadamente até obter levemente para retirar cuidadosamente até obter
adv + v	2689	188	1.759	Quando ferver Quando esfriar quanto baste	não deixe Quando começar não devem Quando levantar
adv + v + prp + v	39	4	12	-	depois vire para dourar muito usado para acompanhar não é para ser novamente misturando sem parar
subtotal	13.588	838	10.024		
art + adj + prp + n	871	59	548	o restante dos ingredientes o fundo da frigideira as laterais da panela	os fundos de alcachofra o doce de leite as laterais com manteiga

art + n + adj	9579	590	6654	a manteiga restante a parte branca a parte interna a parte externa	um creme fofo uma fôrma grande uma peneira grossa a água quente
art + n + adv + adj	686	50	285	a parte mais dura os talos mais macios o recheio já frio a calda ainda quente a parte mais fina	a sopa bem quente o tomate bem picado o chocolate meio amargo a posição mais baixa
art + n + prp + n	21466	1215	7.727	a água do cozimento a casca do camarão a pele do frango a ponta dos dedos o lado da gordura	a batata numa panela a fava de baunilha a salada na hora uma xícara da farinha
subtotal	32.602	1.914	15.214		
n + adj	47214	2031	40.617	fogo baixo vinho branco água fervente filme plástico fogo alto queijo parmesão pimentão vermelho água fria	leite fresco reino branca aspargo saboroso azeite extra banana grande boi moída bolo redonda cabra fresco
n + adj + adj	4199	244	2.924	alface crespa roxa arroz branco cozido coco fresco ralado ervilhas frescas congeladas vinho tinto encorpado	alface americana rasgadas “nova cozinha” francesa coco grande ralado filé mignon cortado peru defumado picado
n + adj + adj + adj	127	7	28	-	pimentão amarelo médio picado pimentão verde médio picado pimentão vermelho grande picado
n + adj + adj + conj + adj	78	2	6	-	camarão miúdo limpo e fervido cebola média picada ou ralada

n + adj + adv + adj + conj + adj	7	0	0	-	presunto cozido bem picado ou moído limões inteiros bem lavados e secos
n + adj + adv + adj + prp + n	6	0	0	-	batata roxa ainda quente no espremedor bolo inglês já molhada com água
n + adj + adv + n	42	0	0	-	prato ralado grosso modo alho picado finamente folhas ovo batido levemente Modo
n + adj + conj + adj	776	34	189	arroz lavado e escorrido frango cozido e desfiado massa lisa e homogênea assadeira untada e enfarinhada	manteiga gelada e picada manteiga restante e polvilhada cenoura cozida e picada
n + adj + n	3831	241	1.510	-	preparo antecipado Ingredientes salsinha picada sal salsinha picada Modo
n + adj + n + n	221	6	25	requeijão cremoso tipo catupiry	leite quente sal Modo limão ralada Acessório forma pimentão verde sal manjerona
n + adj + prp + n	4835	294	1.922	pimenta-do-reino moída na hora gelatina incolor sem sabor superfície polvilhada com farinha azeitonas pretas sem caroço pimenta caiena em pó	alho picados no azeite pó diet de morango assadeira grande com manteiga cenoura média em cubos
n + adj + prp + n + adj	802	45	206	fôrma redonda de aro removível alho-poró cortado em rodela finas queijo parmesão em lascas finas	lata pequena de aspargos escorridos cenoura média em tiras finas gordura extraída das bagas moídas
n + adj + punc + adj + conj + prp + n	10	0	0	-	raiz forte, ralado ou em pó pimentão vermelho, pequeno e sem pele
n + adj + v + prp + n	159	9	42	manteiga gelada cortada em cubos lula limpa cortada em anéis	cebola média cortada em rodela filme plástico encostado no molho
n + adv + adj	1812	112	1.043	chocolate meio amargo tomates bem vermelhos assadeira não untada bolo ainda quente	oliva bem quente sabor mais suave alface bem lavadas

n + conj + n	7735	283	4.531	sal e pimenta-do-reino manteiga ou margarina salsinha e cebolinha	vitaminas e sais Sal e molho reino e noz
n + conj + n + prp + n	1332	35	531	sal e pimenta-do-reino em grãos folhas e talos de salsão sobrecoxas ou coxas de frango	cascas e sementes no creme confeiteiro e canela em pó espinhas e corte-o em pedaços
n + conj + v + adv	1554	95	508	-	ingredientes e misture bem frigideira e aqueça bem casca e lave novamente
n + n	29074	1310	23.749	uva passa cream cheese papel toalha papel alumínio queijo prato	gosto Modo porções Calorias min Rendimento Doces Cozinha
n + n + adj	1592	75	436	uva passa escura uvas passas brancas lingüiça calabresa defumada	gosto queijo parmesão meio tablete dissolvido Acessório forma refratária
n + n + adj + prp + n	315	8	36	uva passa preta sem sementes uvas passas brancas sem sementes	gosto pimenta-do-reino moída na hora limão pimenta-do-reino moída na hora
n + n + punc + n + prp + n + punc	19	1	9	nam pla (molho de peixe)	-
n + prp + n	14520	3134	65.092	azeite de oliva farinha de trigo fermento em pó farinha de rosca essência de baunilha	cm de diâmetro xícara de açúcar litros de água g de chocolate
n + prp + n + adj	79805	863	9.232	creme de leite fresco azeite de oliva extravirgem farinha de trigo peneirada tabletes de fermento biológico	litro de água fervente g de chocolate branco Preparo Numa panela grande
n + prp + n + adj + adj	1268	65	645	tablete de fermento biológico fresco creme de leite fresco batido peito de peru defumado fatiado	sopa de cebolinha verde picada xícara de coco ralado seco g de presunto cru picado

n + prp + n + adj + conj + adj	268	6	29	peito de frango cozido e desfiado dentes de alho amassados e picados	g de ervilhas frescas ou congeladas sopa de amêndoas torradas e picadas
n + prp + n + adv + adj	575	23	248	raspas de chocolate meio amargo folhas de alface bem lavadas tablete de chocolate meio amargo	xícara de cebola bem picada gramas de chocolate meio amargo panela com óleo bem quente
n + prp + n + n	10339	464	5.771	claras em neve firme manteiga em temperatura ambiente ponto de neve firme gemas de ovo batidas	chá de orégano Modo Clafoutis ao kiwi Ingredientes cm de diâmetro Saco consistência de massa Modo
n + prp + n + prp + n	11810	633	6.930	tablete de caldo de galinha fio de azeite de oliva envelope de gelatina em pó claras em ponto de neve	colher de chá de pimenta-do-reino g de chocolate ao leite ml de azeite de oliva bacalhau de molho na véspera
n + prp + n + prp + n + adj	2221	123	954	gelatina em pó sem sabor incolor calda em ponto de fio grosso caixinha de creme de leite light fôrma para torta de aro removível fatias de pão de fôrma branco	colher de chá de páprica picante g de queijo de minas fresco litro de vinagre de vinho branco fermento na metade do leite morno
n + prp + n + prp + n + n	1701	67	311	claras em ponto de neve firme manteiga sem sal em temperatura ambiente claras de ovo em neve firme	tempo de cozimento da massa Rendimento tablete de caldo de carne sal g de farinha de trigo Modo
n + prp + num + n	4952	310	2.790	suco de 1 limão cubos de 2 cm fatias de 1 cm	folhas de 1/2 maço geladeira por 6 horas capacidade para 2 litros
n + prp + v	4734	314	2.818	manteiga para untar saco de confeitar prato de servir forma de abrir	gosto Para polvilhar fogo para aquecer leite até obter
n + prp + v + n	630	37	248	-	batedeira até obter picos manteiga para untar farinha fatias de 1/2 cm

n + punc + n + punc	23964	282	22.198	endro (dill) colher (chá) xícaras (chá) colher (café) copo (requeijão)	farofa . Categoria : manteiga . Fonte : minutos . Dicas : salada . Rendimento :
subtotal	262.527	11.153	195.578		
prp + adj	3473	130	3.127	às colheradas aos pouquinhos ao ponto	de saladas das laterais para média
prp + adv + n	249	17	175	numa só camada com muito cuidado	com bastante manteiga com menos gordura
prp + art + adv + n	139	7	135	de uma só vez em uma só camada em um só pedaço	-
prp + art + n	22661	879	20.555	com um garfo com uma escumadeira Com uma espátula	até o ponto com as sementes com o orégano para uma tigela
prp + art + n + adj	3149	232	2.203	com um batedor manual sem a parte branca até o dia seguinte	com o pão restante em um refratário grande em uma frigideira funda
prp + art + n + adj + prp + adj	13	0	0	-	em uma frigideira grande de fundo sobre uma película fina de líquido com o suco espremido da laranja ¹
prp + art + n + adv	327	25	121	de uma vez só	em uma panela juntamente Com uma faca bem para o recheio não

¹ observe que nesse caso específico, a maioria dos adjetivos da posição final, na verdade, são substantivos que, eventualmente, podem funcionar como adjetivo mesmo no corpus de culinária, mas não nos contextos recuperados por este padrão.

prp + art + n + prp + n + adj	452	31	162	-	com o azeite de oliva restante de uma fôrma de aro removível em uma tigela com água fria
prp + art + n + prp + v	508	42	260	até a hora de servir Com um saco de confeitar até o ponto de enrolar	para um saco de confeitar com a espessura de 1/2 com as mãos até ficar
prp + conj + v	1004	45	847	para que dourem até que engrosse até que murchem	até que estejam até que enfiando até que fique
prp + conj + v + adj	248	23	173	até que fiquem crocantes para que fiquem macias até que estejam cozidos	-
prp + conj + v + adj + conj + adj	11	0	0	-	até que fique macio e mole até que fique opaca e macia
prp + conj + v + art + n (+...) + pron + v + adj	0	6	78	até que enfiando um palito (...) ele saia limpo	-
prp + conj + v + prp + n	65	4	29	até que dobre de volume em que ficou de molho até que esteja al dente em que ficaram de molho	-
prp + n	132770	2193	129.142	em cubos em rodela sem sementes al dente	de forno de vinagre com água em temperatura
prp + n + adj	26237	1149	22.499	No dia seguinte em rodela fina em tiras largas	ao forno pré-aquecido de coco ralado de aro removível
prp + n + adv	1408	87	828	-	dos frutos mais em óleo não na panela novamente
prp + n + adv + adj	899	47	567	em rodela bem fina em lâminas bem finas	de nozes grosseiramente picadas de alho bem picados de salsinha bem picadinhas

prp + n + prp + n	22299	948	16.997	no sentido do comprimento em ponto de neve em direção ao centro na metade do cozimento	pela farinha de trigo para cima numa assadeira em toalha de papel na metade do leite
prp + prp + adv + num + n + prp + n	7	0	0		de pelo menos 3 horas de geladeira por pelo menos 20 minutos em local
prp + v	14417	362	13.687	Em seguida para decorar até ferver para pincelar	até ficarem até começar para derreter de confeitar
prp + v + adj	962	37	794	até ficar homogêneo até ficar macia até ficarem esbranquiçadas	de cheiro verde em conserva cortado para temperar saladas
prp + v + art + n + prp + n	358	17	105	até obter uma textura de areia para retirar o excesso de gordura para eliminar o excesso de óleo até obter o ponto de chantilly	até obter uma calda em ponto de acentuar o sabor do camarão para eliminar as sementes de baunilha
prp + v + conj + v + art + n + prp + n	15	0	0	-	para fritar e corte a banana em tiras para acertar e enfie a pontinha do osso até alisar e acrescente a pasta de avelãs
prp + v + n	1158	65	777	até levantar fervura para dar liga para dar brilho	em conserva sal Para decorar Folhas até obter picos
prp + v + prp + n	802	42	561	até dobrar de volume até ficarem al dente até reduzir pela metade até desgrudar das mãos	para cozinhar em fogo em conserva na salmoura em seguida com salada para amolecer no leite
prp + v + prp + v	712	7	650	sem parar de mexer sem parar de bater	sem tampar até ficar
subtotal	234.343	6.395	214.472		

v + adv	9880	524	8.060	Misture bem mexendo sempre Sirva imediatamente Pique grosseiramente mexa delicadamente	fique bem Sirva bem esfriar completamente ficar levemente
v + adv + adj	1041	76	578	sirva bem quente Desenforme ainda morno Sirva bem gelado	ficar bem dourada estar bem macia estejam bem incorporados
v + adv + art + n	1575	98	533	tampe parcialmente a panela Bata ligeiramente os ovos aperte bem as bordas Lave bem o arroz	Quebre grosseiramente os grãos Misture bem a carne Aqueça bem o azeite
v + art + n	42.907	2.912	29.702	Acerte o sal Abaixe o fogo Escorra o macarrão bata os ovos	Acrescente a cebola Aqueça a manteiga Colocar os camarões passe o caldo
v + art + n + adj + prp + adj	24	0	0	-	despeje a maisena dissolvida no restante forme uma fôrma antiaderente de fundo Coloque o óleo restante na primeira
v + art + n + adj + prp + art + n	254	4	13	-	Bata a manteiga amolecida com o açúcar Junte a farinha peneirada com o fermento
v + art + n + adj + prp + pron	40	0	0		Prepare os crepes restantes da mesma Unte uma forma redonda com esta Separe uma assadeira média de uns forme uma cavidade central em toda
v + art + n + conj + art + n	2247	70	433	acerte o sal e a pimenta-do-reino ajuste o sal e a pimenta	descarte as sementes e os filamentos Elimine a pele e as espinhas Junte a cebola e os tomates

v + art + n + conj + v	5014	234	1565	descarte a espuma que subir	Retire os camarões e reserve separe os gomos e elimine Tampar a panela e cozinhar
v + art + n + conj + v + prp + art + n	351	4	16	-	junte a manteiga e esfarele com a ponta Descasque os camarões e prepare com as cascas
v + art + n + art + num + art	20	3	10	junte os ovos um a um acrescente os ovos um a um frite os bifes um a um	-
v + art + n + pron + art + art	18	2	6	coloque as gemas uma a uma junte as gemas uma a uma	-
v + art + n + prp + art + n	3515	157	921	Faça uma bola com a massa Abra a massa com um rolo Bata a manteiga com o açúcar passe a mistura por uma peneira	Coloque a carne em uma tigela Coloque o óleo em uma frigideira ajuste a acidez com um pouquinho
v + art + n + prp + art + n + adj	576	22	118	Transfira a massa para uma superfície enfarinhada Repita a operação com as porções restantes Abra a massa em uma superfície lisa	Coloque a carne em um refratário médio regue a frigideira com o vinho branco
v + art + n + prp + art + n + adj + prp + n	42	1	4	Coloque a massa sobre uma superfície polvilhada com farinha	-
v + art + n + prp + n	11674	550	4.202	Bata as claras em neve Aqueça um fio de azeite acerte o ponto de sal Derreta o chocolate em banho-maria	coloque o azeite de oliva derreta a manteiga em fogo Acrescente a farinha de trigo
v + art + n + prp + n + adv	174	4	19	-	Aqueça o forno em temperatura bem Bata as claras em neve bem
v + art + n + prp + num + n	728	21	80	Divida a massa em 2 partes deixando uma borda de 1 cm Sove a massa por 5 minutos	ferver o shirataki durante 2 minutos Umedecer a gelatina com 5 colheres

v + art + n + prp + v	558	29	140	Repita a operação até terminar Coloque a água para ferver Repita a operação até acabar	Abra a massa até obter acender um fósforo para flambar
v + art + pron + num + n	51	5	27	mexendo a cada 2 minutos Mexa a cada 2 minutos	-
v + conj + v	4047	282	2.405	tampe e cozinhe Escorra e reserve Desenforme e sirva Unte e enfarinhe	Tampe e leve Escorra e coloque ferver e salgue
v + conj + v + prp + n	855	45	294	-	tampe e leve ao fogo esfriar e coloque na geladeira Escorra e coloque sobre papel
v + n + prp + n	1720	99	742	faça furos na massa alterne camadas de massa espalhe metade do recheio	conserva partidos ao meio decorar Casca de limão decorar Folhas de salsa
v + prp + art + n	8347	584	4969	Passe por uma peneira Regue com o azeite polvilhe com o açúcar Unte com a manteiga	coloque-as em uma panela Temperar com o sal Coloque em um recipiente
v + prp + art + n + adj	1208	75	405	bata com um batedor manual pincele com o ovo batido cubra com um pano limpo	Transfira para uma tigela pequena colocando em uma panela pequena Coloque em um copo longo
v + prp + art + n + prp + n	1539	62	412	misture com as pontas dos dedos regue com um fio de azeite mexendo com uma colher de pau	coloque-as em uma tigela com água coloque-o em uma panela de pressão
v + prp + n	22926	1015	18.231	leve ao microondas Retire do fogo leve ao forno cubra com papel-alumínio	Asse em forno seque com toalha descansar na geladeira
v + prp + n + adj	4361	242	2.910	cozinhe em fogo baixo cubra com filme plástico leve ao fogo alto assar em forno preaquecido	amolecer em água fria crescer em local aquecido Refogar no azeite quente

v + prp + n + adv	324	21	105	leve ao fogo novamente leve ao forno novamente	Retire do fogo depois Frite em óleo bem leve ao forno bem
v + prp + n + conj + n	773	20	505	tempere com sal e pimenta-do-reino Polvilhe com açúcar e canela Salpique com sal e pimenta-do-reino	Retire do fogo e mantenha-os
v + prp + n + n	481	27	191	seque com papel toalha leve ao forno microondas Cubra com papel alumínio	Unte com óleo forminhas cortada em rodela Sal
v + prp + n + prp + adv	1147	30	1001	mexendo de vez em quando regando de vez em quando virando de vez em quando	Deixe na geladeira por cerca Leve ao forno por mais mantenha no fogo por mais
v + prp + n + prp + adv + prp + art + n	47	4	32	mexendo de vez em quando com a colher regando de vez em quando com o caldo	mexendo de vez em quando até a cebola
v + prp + n + prp + adv + prp + art + n + prp + n	16	1	5	regando de vez em quando com o caldo da assadeira	-
v + prp + n + prp + n	2351	116	862	seque com toalha de papel Leve ao fogo em banho-maria deixar de molho na água	leve ao forno em potência Leve ao forno em temperatura
v + prp + n + prp + num + n	1088	72	577	leve ao fogo por 30 minutos deixe no forno por 15 min. leve à geladeira por 4 horas corte em cubos de 2 cm	descansar na geladeira por 30 minutos
v + prp + v	3345	162	2.524	sirva em seguida cozinhe sem tampar leve para assar misture até dissolver	misture até obter mexa até ficar trabalhe até conseguir
v + v	8858	359	7.326	deixe esfriar vá acrescentando deixe crescer espere amornar	deve ficar pode ser deverá ser

v + v + n	1647	68	1.276	deixe tomar gosto Deixe levantar fervura	
v + v + prp + adj	86	9	32	-	é utilizado em diversas deixe esfriar na própria é indicado em grande
subtotal	145.855	8.009	91.231		
total geral	704.129	29.334	541.791		

This document was created with Win2PDF available at <http://www.daneprairie.com>.
The unregistered version of Win2PDF is for evaluation or non-commercial use only.

Anexo VI

As 100 UTEs mais recorrentes por categoria (ordem alfabética)

a água de cozimento	Amasse bem	até que estejam macios
a água do cozimento	anéis de cebola	até que fiquem macias
a água restante	antes de assar	até que fiquem macios
a base da torta	antes de cortar	até surgirem bolhas
a borda da massa	ao marasquino	azeite de oliva
a canela restante	ao rum	bacon cortado em cubos
a casca do camarão	Apague o fogo e acrescente	Baixe o fogo
a casca dos camarões	Aqueça o azeite em uma panela	bastante óleo
a gordura da assadeira	Aqueça o forno em temperatura	bastante açúcar
a manteiga amolecida	as cascas dos camarões	bastante gelo
a mesa enfarinhada	as ervas restantes	bastante manteiga
a metade da altura	as fibras internas	Bata até ficar
a metade da massa	as folhas externas	bata com um batedor manual
a metade da mistura	as folhas verdes	Bata levemente
a metade do açúcar	as laterais da fôrma	bata o creme de leite
a metade do azeite	as partes mais duras	batendo bem
a mistura de azeite	as pontas de aspargos	bem apertado
a mistura de creme	as variedades mais comuns	bem assado
a mistura de fermento	assar em forno preaquecido	bem cremosa
a mistura de manteiga	asse até dourar	bem douradas
a mistura de peixe	assim que amornar	bem dourado
a parte interna	assim que aquecer	bem escorridas
a parte superior	assim que dourar	bem finas
a pele da lingüiça	assim que encorpar	bem fininhas
a pele do peixe	assim que estiver	bem finos
a pele dos pimentões	assim que estiverem	bem fofo
a placa do forno	assim que ferver	bem homogênea
a posição mais baixa	assim que ficar	bem homogêneo
a rúcula restante	assim que iniciar	bem lavada
a salsinha restante	até apurar	bem lavadas
a superfície da carne	até congelar	bem lavado
a superfície da torta	até desgrudar das mãos	bem limpa
a superfície externa	até dobrar de volume	bem lisa
a tampa do pão	até estarem macias	bem macios
a temperatura do forno	até estarem macios	bem misturados
a tigela da batedeira	até ferver	bem miudinho
aceto balsâmico	até ficar cremoso	bem picadas
acrescente à	até ficar transparente	bem picadinha
acrescente ao molho	até ficarem al dente	bem picadinho
acrescente mais	até ficarem macias	bem quentes
água de cozimento	até gratinar	bem rápido
ainda firmes	até murchar	bem seca
ainda morno	até obter	bem seco
ainda quente	até obter o ponto de fio	bem temperada
ainda quentes	até obter picos	bem vermelhos
alecrim seco	até obter ponto	berinjela em rodela
amarrado de ervas	até que estejam firmes	biscoito champagne

broto de feijão	com um pincel umedecido	deixe esfriar
cabeças de alho	com uma escumadeira	Deixe ferver
café expresso	Com uma espátula	deixe fritar
capacidade para 200 ml	como aperitivo	deixe hidratar
carne de porco	como condimento	Deixe marinar
casca de 1 limão	como decoração	deixe no freezer
casca do limão	como um bife grande	Deixe repousar
cebola média em rodela	concentrado de tomate	demaís ingredientes
cebola média ralada	contra filé	dente de alho picadinho
cebolinha verde picada	cortada em rodela	dentes de alho em lâminas
cebolinhas verdes	cortada em tiras finas	dentes de alho inteiros
cerca de 1 5 kg	cortadas ao meio	depois de 15 minutos
cerca de 12 cm	cortadas em cubinhos	depois de 3 minutos
cerca de 12 minutos	cortadas em cubos	depois de esfriar
cerca de 15 minutos	cortadas em fatias	derreta a manteiga em fogo
cerca de 2 cm	cortadas em gomos	descascadas e cortadas
cerca de 2 horas	cortadas em pedaços	descascadas e picadas
cerca de 20 cm	cortadas em rodela	Desenforme e sirva
cerca de 200 g	cortadas em tiras	desenforme e sirva
cerca de 25 cm	cortado ao meio	deve estar
cerca de 25 minutos	cortado em cubinhos	dissolvida na água
cerca de 250 g	cortado em cubos	dissolvida no leite
cerca de 30 minutos	cortado em cubos pequenos	dissolvido na água
cerca de 4 minutos	cortado em fatias	dissolvido no leite
cerca de 5 cm	cortado em pedaços	dobre de volume
cerca de 5 minutos	cortado em rodela	doce de leite
cerca de 6 minutos	cortado em tiras	dourada e crocante
cerca de 8 cm	cortado em tiras finas	dourar de maneira
chocolate em pó	cortado em tirinhas	dourar levemente
coentro em grão	cortados ao meio	em anéis finos
Colocar numa travessa	cortados em 4	em assadeira untada
Coloque em um liquidificador	cortados em cubos	em azeite
coloque em um recipiente	cortados em fatias	em banho maria
coloque em uma panela	cortados em gomos	em bastões
Coloque em uma tigela	cortados em lâminas	em cima
Coloque na forma	cortados em pedaços	em cubos grandes
coloque no liquidificador	cortados em rodela	em excesso
coloque o numa tigela	cortados em tiras	em fio
coloque por cima	corte em pedaços	em flocos
com água suficiente	cozido e desfiado	em fogo alto
Com a ajuda	cozinhar em fogo brando	Em fogo baixo
com aro	Cozinhe em fogo alto	em fogo mínimo
com as costas	Cozinhe em fogo baixo	em forno quente
Com o óleo reservado	creme de cebola	em forno regular
com o óleo restante	creme de leite azedo	em lâminas finas
Com o óleo restante	crescer e dourar	em panela tampada
com o bico	cubra com a água	em pedaços médios
com os dedos	cubra e deixe	em pedaços menores
com pele	decore com fatias	em pedaços regulares
com recheio	decore com folhas	em quadrados
Com um batedor	deixe assar	em rodela grossas
Com um garfo	Deixe cozinhar	em rodelinhas finas

em rolinhos	grosseiramente moída	muito mole
Em seguida	já assada	muito usado
em seguida	já fria	muito utilizado
em tiras bem finas	já frio	não muito grandes
em um lugar	líquido de cozimento	não muito quente
em uma espessura uniforme	líquido do cozimento	não untada
em uma panela com água fria	lado grosso do ralador	na diagonal
enfiando um palito no centro	lascas de coco	na gordura do bacon
enquanto isso	lata de atum	na horizontal
ervilha fresca	lata de milho	na máquina
escorra a água e reserve	laterais da fôrma	na metade do tempo
escorra bem	laterais da panela	na ordem
escorrer numa peneira	laterais de uma assadeira	na pressão
esfriar completamente	laterais de uma fôrma	na salmoura
especial para diabéticos	lavado e escorrido	no centro dos pratos
espere ferver	levar à geladeira	no fogo alto
espremedor de batatas	Levar ao forno médio	No liquidificador
excesso de sal	levar ao forno médio	no ponto
fôrma para bolo	leve ca fervura	no preparo de pratos
fatias de 1/2 cm	leve a ao fogo	no sentido longitudinal
feche bem	Leve ao fogo	no tempero
fervor em fogo alto	leve ao forno	nozes moídas
filés de anchovas	levemente amassados	o açúcar restante
filés de robalo	levemente batidos	o azeite reservado
filés de salmão	levemente dourada	o creme reservado
final do cozimento	levemente dourado	o excesso de óleo
fio negro do dorso	levemente encorpado	o frango reservado
fogo médio	levemente untado	o fundo da fôrma
folhas de couve	licor de cassis	o fundo da forma
folhinhas de hortelã	ligeiramente batido	o fundo da frigideira
forma de buraco	ligeiramente dourados	o líquido de cozimento
forma refratária	ligeiramente espesso	o líquido do cozimento
forminhas com capacidade	lisa e elástica	o leite restante
forno em temperatura alta	lisa e homogênea	o molho da carne
forno médio	lisa e macia	o molho de cozimento
forno preaquecido	maças verdes	o molho do tempero
forrada com papel manteiga	maço de ervas	o molho restante
fundo da assadeira	maduros e firmes	o ovo restante
fundo da fôrma	massa em excesso	o pão restante
fundo da forma	mexa bem	o pau de canela
fundo da frigideira	Mexa bem	o recheio já frio
fundo da massa	mexa com cuidado	o refogado de cebola
fundo da panela	microondas em potência alta	o refogado de legumes
fundo de um refratário	misturar bem	o restante da água
fundo de uma assadeira	Misture delicadamente	o restante da cebola
fundo de uma fôrma	misture novamente	o restante da manteiga
fundo de uma panela	moída grosseiramente	o restante do caldo
fundos de alcachofra	moída na hora	óleo de gergelim torrado
gelatina em pó sem sabor incolor	Molho branco	óleo extravirgem de oliva
gemas batidas	molho do cozimento	os filamentos brancos
gomos finos	molho homogêneo	os ingredientes da massa
gordura do bacon	muito líquido	os ovos restantes

os talos grossos	quente ou frio	tiras bem finas
os talos mais grossos	ralada grosso	tiras de massa
pão amanhecido	ralado grosso	tomates médios sem pele
palmito em conserva	ralo grosso	Transfira a massa para uma
panela semitampada	ramo de hortelã	superfície enfarinhada
pano limpo	Refogar na manteiga	um creme bem fofo
Para pincelar	Refogar no azeite	um creme bem liso
Para untar	refogue até ficar	um creme encorpado
parte escura	refogue em fogo alto	um creme espesso
pé de alface	refogue em fogo baixo	um creme fofo
peito de frango cozido	regue com azeite	um creme levemente encorpado
peito de peru defumado	regue com o azeite de oliva	um molho grosso
peitos de pato	reserve na geladeira	um purê homogêneo
pele dos pimentões	restante da água	um tablete dissolvido
pelo comprimento	restante da farinha	uma calda caramelada
pelo espremedor	restante da manteiga	uma calda grossa
pelo processador	restante da massa	uma camada de fatias
peneira de malha fina	restante do óleo	uma consistência cremosa
peneira grossa	restante do açúcar	uma crosta dourada
picada em pedaços pequenos	restante do azeite	uma espessura uniforme
picada finamente	restante do leite	uma farofa úmida
picada grosseiramente	restante dos ingredientes	uma gemada clara
picadas finamente	retângulos de massa	uma massa bem lisa
picadas grosseiramente	Retire do fogo	uma massa lisa
picado em cubos	retire do fogo	uma mistura bem cremosa
picado em pedaços	rodela de abacaxi	uma mistura bem lisa
picado finamente	rodela de limão	uma mistura clara
picado grosseiramente	rodela fina	uma parte da casca
picados grosseiramente	salgue levemente	uma pasta grossa
picos firmes	salteando de vez em quando	untada com óleo
pimentão vermelho sem sementes	seleta de legumes	untada com azeite
pimenta do reino em grão	sem as sementes	untada com manteiga
Pincele com a gema	sem encostar	untada e enfarinhada
pique as finamente	sem filamentos	untada e polvilhada
Pique bem	sem lavar	untadas com manteiga
polvilhada com farinha	sementes de erva doce	vá acrescentando
por completo	sirva bem quente	Vá acrescentando
por um espremedor	Sirva com arroz	vinagre de vinho branco
por unidade	sirva com o molho	vinho tinto seco
pote de iogurte	Sirva com o molho	volte ao fogo
própria para cozimento	Sirva com pão	xícara de café
própria para microondas	Sirva como	xícara de chá
próprio para microondas	superfície polvilhada	xícaras de chá
quando a carne estiver	tampe a panela e deixe	
quando a manteiga estiver	Tampe e cozinhe em fogo	
quando começar	tampe e cozinhe em fogo	
quando perfumar	temperada com sal	
quanto baste	temperados com sal	
quase pronto	Temperar com o sal	
queijo gouda	tempere os com sal	
queijo pecorino	tigela da batedeira	
quente ou fria	tigela funda	

This document was created with Win2PDF available at <http://www.daneprairie.com>.
The unregistered version of Win2PDF is for evaluation or non-commercial use only.

Anexo VII

Exemplo de visualização inicial de verbete no DOC-Trad

Home	STU (Specialized Translation Unit):	freq. SL corpus:
Tradução Culinária	pimenta-do-reino moída na hora	213	
Como usar o DOC-Trad	Translation:	freq. TL corpus:	
Consultar DOC-Trad (por palavra)	freshly ground black pepper	2.446	
	freshly ground pepper	249	
Consultar DOC-Trad (por palavra-chave)	Translator's notes:		our files other sources: web videos Google Image
Consultar DOC-Trad (por área)	Although in English the adjective “black” is seldom omitted, in Portuguese, the corresponding adjective, “preta”, is used mostly when there are other types of pepper being used in the recipe.		
Consultar CorTrad (corpus paralelo)	Examples (original Portuguese → translated English):		
Consultar CorTec (corpus comparável)			
Pesquisar no Google:			
Usar o KWIC Google			
Consultas externas			

Esfregue o pernil com azeite, sal e pimenta-do-reino moída na hora	Rub fresh ham with olive oil, salt and freshly ground black pepper
Sal e pimenta-do-reino (moída na hora , de preferência) a gosto	Salt and black pepper (preferably freshly ground), to taste
Tempere os camarões com o sal e a pimenta moída na hora	Season shrimps with salt and freshly ground pepper

→ [show more bilingual contexts from CorTrad parallel corpus](#) (up to 10)

→ [show more contexts from CorTec comparable corpus – English](#) (up to 10)

Fórum de dúvidas	Concurrent STUs in SL:	freq. in corpus:
	pimenta moída na hora	6
Contato	pimenta-do-reino preta moída na hora	6
	Related STUs in TL:	
	<u>pepper₁</u> ; <u>black pepper</u> ; <u>white pepper</u> ; <u>pink pepper</u> ; <u>green peppercorns</u> ; <u>salt and pepper to taste</u> ; <u>mixed peppercorns</u> ; <u>coarsely ground PEPPER₁</u> ; <u>pepper shaker</u> ; <u>pepper mill</u> ; <u>sprinkle with salt and pepper</u> ; <u>freshly ground PEPPER₁</u> ; <u>cracked PEPPER₁</u>	
	links to the hypertree:	<u>ingredient</u> <u>process</u>

This document was created with Win2PDF available at <http://www.daneprairie.com>.
The unregistered version of Win2PDF is for evaluation or non-commercial use only.

*Eu sou é eu mesmo.
Divêrjo de todo mundo...
Eu quase que nada não sei,
mas desconfio de muita coisa.*

João Guimarães Rosa
Grande Sertão: Veredas, ∞

This document was created with Win2PDF available at <http://www.daneprairie.com>.
The unregistered version of Win2PDF is for evaluation or non-commercial use only.