

# LINGÜÍSTICA DE CORPUS E TRADUÇÃO TÉCNICA – RELATO DA MONTAGEM DE UM CORPUS MULTIVARIETAL DE CULINÁRIA

**Stella E. O. Tagnin\***  
**Elisa Duarte Teixeira\*\***

RESUMO: A Lingüística de Corpus, enquanto metodologia, tem se mostrado de extrema valia para a Tradução Técnica, seja para a prática tradutória em si, seja para o desenvolvimento de suas áreas afins como, por exemplo, a Terminologia. No presente artigo, relatamos a construção de um corpus multivarietal comparável de receitas em português brasileiro/europeu e inglês americano/britânico. A motivação para tal pesquisa foi a presença cada vez maior de obras da área editadas em português europeu no mercado editorial brasileiro. A exploração do corpus coletado revelou que as diferenças lexicais, sintáticas e culturais entre as variantes do português, comparadas às das variantes do inglês, são mais expressivas, justificando traduções na variante brasileira e obras terminológicas que explicitem essas diferenças.

UNITERMOS: Lingüística de Corpus; Tradução Técnica; Critérios de construção de corpora; Estudos Contrastivos; Culinária

*ABSTRACT: Corpus Linguistics, as a methodology, has been shown to be extremely valuable for Technical Translation, both for translation practice and terminological research. This article reports on the construction of a multivariant corpus of cooking recipes in Brazilian and European Portuguese and American and British English. Our research was motivated by the ever-growing number of European*

---

\* Universidade de São Paulo, Brasil.

\*\* Aluna de mestrado em Língua Inglesa pela Universidade de São Paulo ([elisadut@usp.br](mailto:elisadut@usp.br)), bolsista Capes.

*Portuguese cookbooks sold in the Brazilian market. The investigation of our corpus has revealed that the lexical, syntactic and cultural differences between the Portuguese variants are more expressive than those between the English variants. We claim that this would justify Brazilian translations of these books and terminological reference works which make these differences explicit.*

*KEYWORDS: Corpus Linguistics, technical translation, corpus building: criteria and organization, contrastive studies, Cooking.*

### **I. Tradução técnica e Terminologia – a situação da Culinária no Brasil**

A Culinária, além de ser um saber que permeia várias áreas do conhecimento (como a literatura, a história, a nutrição, a agricultura, a química e a física), carrega consigo marcas indelévels da cultura de um povo, região, e/ou país. No entanto, tanto o mercado editorial brasileiro quanto os terminólogos e estudiosos da área parecem não ter se dado conta da ligação profunda entre a cultura de um país e seu universo culinário. Isso pode ser atestado, por exemplo, pela presença aparentemente grande de edições portuguesas de livros de receitas e obras de referência (como glossários e guias) nas livrarias brasileiras. Até mesmo as publicações terminológicas brasileiras na área, que são escassas, parecem não fazer distinção entre as variantes brasileira e européia. Veja-se, por exemplo, o *Dicionário-Almanaque de Comensais & Bebês*, de Cláudio Fornari (2001), que traz, à página 27, estas duas definições, na seqüência:

**Alourar.** Dar aos alimentos uma cor dourada, tostando-os no fogo ou no forno, inclusive pincelando-os previamente com uma gema (no caso de massa). Corar.

**Alperche.** Tipo de damasco grande, com cheiro semelhante ao do pêssego.

Embora tais definições possam estar “corretas” (diferentemente da definição que o autor dá para **Allspice**.<sup>1</sup> “*Mistura de cravo, gengibre, pimenta-do-reino, noz-moscada e canela, podendo receber outros temperos. V. tb. Cinq épices.*”), do ponto de vista da tradução técnica, a presença desses termos seria inconcebível, tanto numa tradução de receita brasileira quanto num glossário técnico da área. Isso porque, ainda que esses termos façam parte do inventário da língua geral, não são comumente usados na linguagem técnica corrente da variante brasileira. Como salientou Roman Jakobson em seu texto clássico “Aspectos lingüísticos da tradução” (1979): “*As línguas diferem essencialmente naquilo que devem expressar, e não naquilo que podem expressar*”. Essa é uma preocupação central da tradução, especialmente da tradução técnica – a naturalidade: como as coisas, de fato, são ditas numa dada área de uma dada língua ou variante lingüística.

Nesse sentido, a Lingüística de Corpus tem exercido um papel incontestável na evolução dos estudos tradutológicos e terminológicos, por insistir na noção de linguagem enquanto objeto de pesquisa cujos fenômenos só podem ser observados em textos autênticos, produzidos naturalmente, sejam eles escritos ou falados (Teubert, 2001). Dicionários bilíngües, de modo geral, costumam apresentar apenas uma lista de equivalentes, de pouca ou nenhuma ajuda ao tradutor, pois não fornecem informações que permitam uma adequação dos termos a um contexto específico (Váradi e Kiss, 2001). Além disso, a terminologia tradicional costuma desprezar as ocorrências multi-palavras – ou *lexias complexas*, ou *colocações*<sup>2</sup> –, via de regra, as que apresentam maior dificuldade na tradução.

O primeiro dicionário de língua geral a dar um tratamento mais adequado a essas unidades multi-palavras fazendo uso de dados provenientes de corpora eletrônicos foi o *Collins Cobuild English Dictionary*, de língua inglesa, editado em 1987, sob a su-

<sup>1</sup> Pimenta-da-jamaica

<sup>2</sup> Em inglês, “collocations”. Termo introduzido por Firth, é uma “*unidade lingüística que se caracteriza pela combinação consagrada de seus elementos*” (Tagnin, 1989, p. 81), por exemplo, “pimenta-do-reino moída na hora” ou “claras em neve”

pervisão do lingüista John Sinclair. O uso de corpora vem se difundindo de tal maneira desde então que “fazer lexicografia sem usar corpora eletrônicos é algo praticamente impensável nos dias de hoje”,<sup>3</sup> conforme afirma Willem Meijs (1996, p. 100). E isso se aplica também à terminologia. De acordo com Baker (1995, p. 223), uma especialista na área dos Estudos Tradutológicos: “a terminologia baseada em corpus, atualmente, não é a exceção, mas sim a regra”.<sup>4</sup> A terminologia no Brasil, nesse sentido, ainda se encontra, aparentemente, mais voltada para a ‘exceção’ do que para a ‘regra’.

Contudo, além dos dicionários, o tradutor da área também pode apoiar-se nas informações dispostas em “guias”, os quais, muitas vezes, estão centrados em subáreas específicas da culinária, tais como as técnicas de cozinha (preparação de carnes, doces, licores; uso de microondas; etc.) ou os ingredientes (temperos; frutas; cereais; etc.). Esse tipo de material é mais difícil de se encontrar, pois, muitas vezes, são anexos, separatas, brochuras que acompanham revistas ou produtos em supermercados, etc. Mas alguns, mais bem cuidados, chegam às livrarias, como veremos a seguir.

O livro *Ervas e especiarias na cozinha* (Junqueira, 1980), por exemplo, traz uma minuciosa descrição de cerca de 70 variedades de ervas, especiarias e temperos, com: a) nome científico; b) ilustração; c) termos equivalentes em espanhol, francês, italiano, inglês e alemão;<sup>5</sup> d) um breve histórico do uso do tempero; e e) como comprar e como usar, entre outros. Em seguida há por volta de cem receitas em que esses são empregados como ingredientes. Seria uma obra muito útil para consulta, caso seu índice remissivo remetesse não só às informações completas de cada tempero, mas também às receitas em que são empregados,

<sup>3</sup> “*lexicography without computerized corpus data is practically unthinkable nowadays*”

<sup>4</sup> “*corpus-based terminology compilation is now the norm rather than the exception*”

<sup>5</sup> Maria Lúcia Gomensoro, autora do *Pequeno Dicionário de Gastronomia* (1999), baseou-se nesse livro de Junqueira para montar seus verbetes “multilíngües” de ervas e especiarias, muitas vezes copiando *ipsis litteris* a informação fornecida.

para que o tradutor verificasse o uso do termo no contexto culinário e suas possíveis colocações (por exemplo, “pimenta-do-reino *moída na hora*” ou “canela *em pau*”).

Além dos dicionários, não poderíamos deixar de mencionar – e também comentar – aquelas obras que, embora não sejam de referência, podem trazer muitas informações relevantes para o tradutor da área.

Uma delas é a intitulada *Ingredientes* (Werle & Cox, 2000), no formato 21x28 cm, com 384 páginas e incontáveis fotos, uma para cada ingrediente relacionado na obra. Ou seja, um oásis para o tradutor ou especialista/pesquisador da área! No entanto, a miragem logo se desfaz: apesar de um brasileiro assinar a tradução (conforme soubemos depois), o que se vê na obra são páginas e mais páginas de termos em português... europeu! *Lemon grass*, da primeira página das “Ervas”, por exemplo, virou “Erva-de-limão”, em vez do nosso “capim-cidreira / capim-limão / capim-santo”...

Essa mesma erva foi o termômetro a avisar que *O Guia das Especiarias. Guia do conhecedor* (Craze, 1998), de capa dura, muitas ilustrações e 192 páginas, padecia do mesmo problema – um certo hermetismo para o leitor brasileiro (aqui a erva mencionada acima recebe outro nome ainda: *chá-do-príncipe*; essa equivalência só foi percebida pela foto e pelo nome científico coincidentes). A capa, a estante da livraria, a folha de catalogação da obra (não consta local de publicação), nada informa ao consumidor desavisado que a publicação é em português europeu e que, muito provavelmente, em vez de adquirir um guia muito útil para sua biblioteca, “Essencial para usar na Culinária, Cosmética e Medicina” (dizeres da capa do livro), ele está comprando um guia de curiosidades, que lhe causará muito estranhamento: grãos-do-paráiso, galanga-de-java, açafroa, cravinho, ênula-campana, murta, papoila-da-índia, quássia, calicanto, cubeba e alforva são algumas das especiarias elencadas no livro.

Para ilustrar melhor essa questão das escolhas lexicais feitas por culturas distintas, transcrevemos abaixo uma receita escrita originalmente em português europeu.<sup>6</sup> Qual não seria o

<sup>6</sup> Receita retirada do *site* português: [www.gastronomia.oninet.net](http://www.gastronomia.oninet.net), em maio de 2003.

espanto do entusiasmado *chef-de-fim-de-semana* brasileiro ao se deparar com o seguinte texto?:

**Tarte de Sapateira com Sultanas**

**Ingredientes:**

Uma sapateira      4 gemas  
 200 gr de massa folhada   50 gr de *pickles*  
 50 gr de sultanas      15 gr de queijo parmesão  
 ½ chávena de maionese   Sal e pimenta q.b.  
 ½ chávena de natas

**Preparação:**

Coze-se a sapateira em água temperada com sal. Depois de arrefecer abre-se e retira-se-lhe a cauda e a tripa, aproveitando o líquido, a carne e o coral. Com um garfo desfaz-se tudo e mistura-se a maionese, as natas, as gemas, as sultanas, os *pickles* picados. Tempera-se com sal e pimenta.

Estende-se a massa folhada e forra-se com ela uma forma de tarte. Deita-se o recheio na forma, polvilha-se com queijo ralado.

Leva-se a cozer 40 minutos, em forno médio. Serve-se quente ou fria, acompanhada com salada.

**Figura 1:** Receita originalmente escrita em português europeu

Como é possível perceber, as diferenças lexicais podem, muitas vezes, impossibilitar a execução da receita – que é, a princípio, a finalidade única de se escrever e/ou ler uma receita. Assim, o que nos preocupa sobremaneira é o fato, impercebido por muitos leitores e desconsiderado por muitas editoras, de essas obras não trazerem qualquer alerta quanto à disparidade entre o público-alvo para o qual foram produzidas originalmente e aquele que as está consumindo aqui, em solo brasileiro – fato esse capaz de destituir o texto de sua função retórica original. Apesar de haver poucas iniciativas dentro dos Estudos Tradutológicos quanto ao estudo contrastivo da retórica textual, autores como Kaplan (1966) e Hinds (1983) (*apud* Colina, 1997) demonstraram que as características textuais variam não só de língua para língua, mas também de cultura para cultura.

Ora, o que é feito, então, do mercado brasileiro de tradução técnica? Será que o mercado editorial brasileiro, a exemplo do que acontece com os mercados americano e britânico, sentenciou que a existência de uma língua comum justifica a intercambialidade das obras dessa área nas duas variantes do português? Será que essa intercambialidade se dá com a mesma facilidade nas variantes das duas línguas? Como fica, por sua vez, a questão da aplicabilidade das informações veiculadas nesses livros se o consumidor brasileiro, desprevenido, leva para casa um livro com receitas que ele não compreende – sejam os ingredientes, sejam os processos – e não pode portanto executá-las?

## II. Estudos Lingüísticos e Culinária – Relato da construção do corpus ‘multivarietal’

### 1. Estudos contrastivos

O estudo comparativo das variantes britânica e norte-americana do inglês é um tema que há décadas vem merecendo a atenção de lingüistas e teóricos da tradução (Johansson, 1980, p. 85). O mesmo não ocorre, no entanto, com o português e suas variantes brasileira (PB) e européia (PE). Wittman, Pêgo e Santos (1995) assim comentam:<sup>7</sup>

Dicionários contrastivos como o de Mauro Vilar (1989),<sup>8</sup> trabalho rico de informações e elaborado com muito rigor, e até mesmo o dicionário humorístico de Mário Prata (1993),<sup>9</sup> constituem os únicos registros voltados especificamente à observação dos contrastes entre PE e PB, além do trabalho de Tereza Biderman<sup>10</sup> (...).

<sup>7</sup> Além dos trabalhos mencionados abaixo pelas autoras, existe também o *Dicionário lusitano-brasileiro*, de Eno Teodoro Wanke, edição de Ediouro, com 176 páginas, esgotada no momento.

<sup>8</sup> VILAR, M. (1989) *Dicionário contrastivo luso-brasileiro*. Rio de Janeiro: Ed. Guanabara.

<sup>9</sup> PRATA, M. (1993) *Schifaizfavoire. Dicionário de Português*. São Paulo: Ed. Globo S.A.

<sup>10</sup> BIDERMAN, M.T. (1994) *Vocabulário fundamental: Cultura e sociedade*. Araraquara: UNESP.

Segundo as autoras, “não desmerecendo o inegável mérito”, o trabalho de Biderman por elas citado sofre de um problema: apoiar-se em corpora não comparáveis. De qualquer forma, no que tange ao assunto que nos interessa – a linguagem técnica da culinária – dois artigos de Biderman (1992 e 1996) apresentam evidências da importância e especificidade lingüístico-cultural dessa área no léxico de uma língua.

No artigo de 1992, a autora lista exemplos de palavras que, conforme mostraram suas pesquisas, deveriam constar num índice de vocabulário fundamental do português brasileiro, por nomearem referentes específicos da realidade brasileira ou por usarem um significante diferente daquele usado em Portugal para o mesmo referente e/ou conceito. Das 120 palavras listadas, aproximadamente 35% têm relação direta com o universo de discurso da culinária. De modo semelhante, ao listar a contrapartida portuguesa, pelo menos 15% dos vocábulos pertence ao universo de discurso da culinária.

Em seu artigo de 1996 (p. 41), ao discutir “Um vocabulário fundamental para o português do Brasil”, Biderman destaca, mais uma vez, as diferenças entre as variantes brasileira e europeia do português. É possível perceber, novamente, a presença significativa de vocábulos relacionados à culinária, fato que merece o comentário da autora: “(...) em outros domínios culturais, como a culinária e a música, por exemplo, encontraremos vocabulário ligeiramente diferenciado em virtude das especificidades de cada uma das culturas de língua portuguesa” (p. 43).

Ainda que representem um avanço dos estudos contrastivos do português para a variante brasileira, não figura dentre as tentativas isoladas de autores como os mencionados acima qualquer trabalho voltado especificamente para áreas técnicas do conhecimento, estudos esses de grande valia para a teoria e a prática da tradução.

## 2. Lingüística de Corpus

Estudos lingüísticos baseados em corpora – coleções de textos, palavras, frases, trechos, diálogos, etc. – são realizados desde o final de 1800, por autores como Pryer (1889) e Stern



(1924), os quais estudaram a aquisição da linguagem em diários, descrevendo cuidadosamente todas as locuções da criança observada. Ou mesmo Kāding (1987) que, com a ajuda de uma numerosa equipe, montou e analisou manualmente um corpus de 11 milhões de palavras para estudar o padrão de distribuição de letras no alemão (McEnery e Wilson, 1997).

No entanto, a Lingüística de Corpus (LC), atualmente, é considerada uma metodologia de pesquisa específica em que a exploração da linguagem é feita por meio de evidências, extraídas por meio do uso de ferramentas computacionais, de um corpus de linguagem natural e autêntica criteriosamente reunido e disponível eletronicamente. A LC vem se destacando cada vez mais na lingüística moderna, apesar de ter sido relegada a segundo plano nos anos 60 e 70, quando as teorias da Gramática Gerativa de Chomsky passaram a dominar os estudos lingüísticos.

Segundo Baker (1995, p. 225), a palavra *corpus* é hoje usada para designar:

(...) 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 for (...).

A montagem de um corpus, no entanto, não é tarefa corriqueira. Os critérios empregados são, de fato, um fator crucial para o sucesso das pesquisas que o tomarão como base, pois o corpus oferece ao pesquisador apenas aquilo que nele foi inserido. Vários autores, como Quirk (1992), Biber (1993), Baker (1995) e McEnery & Wilson (1997), têm se dedicado ao estudo desses critérios. Segundo afirmam Atkins *et al.* (1992), num dos primeiros estudos a tratar a fundo esse assunto sob a perspectiva da Lingüística de Corpus, o processo de criação de um corpus envolve cinco estágios principais:

- A. Planejamento** (*specification and design*): é preciso estabelecer de antemão que tipo de corpus se pretende construir, levando-se em conta fatores como:
- quem usará o corpus e para que?
  - qual o tamanho – de cada texto e do corpus final – e qual o tipo de texto a ser incluído (amostras ou textos integrais?, falados ou escritos?);
  - quais as variedades lingüísticas representadas (quanto ao lugar: sintópico / diatópico?; à data: sincrônico / diacrônico?; à camada social: sistrático / diastrático?; aos atos de fala: sinfásico / diafásico?);
  - qual o nível aproximado de codificação (será etiquetado? como?)
- B. Suporte técnico** (*hardware and software*): a partir do planejamento pode-se estimar a capacidade física necessária dos computadores que irão armazenar e processar o corpus, bem como criar ou selecionar os programas que serão usados para codificá-lo e, posteriormente, explorá-lo: concordanceadores, mecanismos de busca, programas estatísticos, etc.
- C. Coleta e preparação dos dados** (*data capture and mark-up*): o próximo passo é coletar os textos e organizá-los, de forma clara, em meio eletrônico. Nessa etapa são feitas também, conforme a escolha do pesquisador, as anotações (tags) e a etiquetagem de suporte ao cabeçalho (header). Esse processo será mais bem detalhado quando da descrição da construção do nosso corpus (item 2.3).
- D. Processamento do corpus** (*corpus processing*): a penúltima fase da construção consiste na análise quantitativa e qualitativa do corpus e dos resultados obtidos, fazendo uso de todas as ferramentas computacionais escolhidas. Nesse momento é possível verificar as lacunas, falhas e/ou incongruências de todo o processo para então proceder à fase final da construção do corpus.
- E. Aumento e monitoração do corpus** (*corpus growth and feedback*): nessa fase corrigem-se as imperfeições detectadas na fase anterior, aumentando e/ou reorganizando o corpus. É o momento também de colocá-lo à disposição de um público especializado para que possa usá-lo em suas pesquisas e, assim, oferecer comentários ou sugestões de parâmetros de alteração. Essa fase pode repetir-se indefinidamente, dependendo

do propósito do corpus e das necessidades dos pesquisadores envolvidos.

A seguir, descreveremos a primeira etapa de montagem do nosso Corpus de Culinária Quadricultural, utilizando-nos das etapas sugeridas pelos autores acima.

## 2.1 Planejamento do corpus

Conforme já foi dito, nossa hipótese de trabalho constitui-se na verificação de que as diferenças entre o português do Brasil e de Portugal são lingüística e culturalmente mais expressivas do que entre o inglês americano e o britânico. Partindo, pois, dessa hipótese, o primeiro passo para o planejamento do corpus foi determinar o que seria um corpus minimamente representativo dessas línguas em se tratando de receitas culinárias. Foram estabelecidos os seguintes critérios de coleta para orientar a construção do corpus:

### 2.1.1 Quanto ao tipo de corpus

Nossa intenção era montar um “corpus quadricultural comparável”, ou seja, para cada receita incluída no corpus teríamos um “equivalente” em cada uma das quatro variantes em discussão. Para tanto, a semelhança entre as receitas deveria basear-se, sempre que possível, no nome do prato, nos ingredientes que as compunham e no tipo de processo empregado. Por exemplo:

- **Português brasileiro:** “Torradinhas de alho”    *européu:* “Pão com alho”
- **Inglês americano:** “Garlic bread”    *britânico:* “Garlic bread”

Isso, muitas vezes, foi problemático (vide item 2.2.3), já que muitos ingredientes, tipos de prato e até mesmo processos não são empregados ou não são comuns nas quatro culturas.

### 2.1.2 Seleção de textos

A seleção dos textos seguiu os seguintes critérios:

TRADTERM, **10**, 2004, p. 313-358

- **somente receitas** (glossários e artigos de culinária, p. ex., não entraram);
- receitas em que **ingredientes** e **modo de fazer** constituíssem **seções distintas**;
- **textos originais** (não traduzidos) em cada uma das línguas escolhidas. Criamos, assim, siglas de três letras maiúsculas para representar as variantes:
  - Português Original do Brasil = **POB**
  - Português Original de Portugal = **POP**
  - Inglês Original Americano = **IOA**
  - Inglês Original Britânico = **IOB**;
- textos com **boa qualidade gramatical e ortográfica** (textos com muitos erros gramaticais, sem acentos e cedilhas – no caso do português –, por exemplo, foram descartados). Procuramos, na medida do possível, recolher receitas de fontes “idôneas” – *sites* mais reconhecidos, em que fosse previsível um trabalho prévio de edição para publicação na rede. Muitas vezes, no entanto, não tivemos escolha e acabamos por incluir receitas de *sites* pessoais para poder garantir a comparabilidade almejada. Ainda assim, os critérios de boa qualidade do texto foram observados.
- textos cujo processo incluísse uso de forno microondas, máquina de pão (*breadmachine*) ou churrasqueira, por exemplo, não foram incluídos, pois acreditamos que a presença desses aparatos simplificava a descrição dos processos e, no caso da churrasqueira, **não fazia parte do “dia-a-dia” da culinária caseira**;
- **linguagem cotidiana** (não coletamos receitas assinadas por grandes *chefs*, com ingredientes raros/exóticos ou linguagem rebuscada/subjetiva. Receitas de culinaristas, no entanto, foram incluídas);
- **linguagem geral** (evitamos, dentro do possível, qualquer receita de cunho marcadamente regional ou étnico, receitas de dietas especiais – como receitas *light*, para diabéticos ou pessoas com intolerância ao leite – ou *vegan*<sup>11</sup>);

<sup>11</sup> Enquanto o vegetariano elimina a carne da sua dieta, seja ela branca ou vermelha, o *vegan* é mais radical: além da carne, não utiliza nenhum produto de origem animal - incluindo o leite e seus derivados, ovos,

- **textos disponíveis eletronicamente** (optamos por não escanear ou digitar receitas, por mais convenientes/adequadas que fossem. Só coletamos textos de revistas e livros se os mesmos estivessem disponíveis na Internet).

### 2.1.3 Quanto às categorias (subcorpora)

Após pesquisarmos vários sites de receitas, percebemos que a forma como eles separam suas receitas é variada. Muitas vezes, por exemplo, categorias diferentes de receitas são misturadas, como as etapas de uma refeição (“entrada, prato principal, sobremesa” etc.) junto de categorias que não se encaixam aí (“salgadinhos, bolos, sanduíches”, etc.). Isso levou-nos a criar uma categorização também mista (vide **Figura 2**). Chegamos a essas 11 categorias, que convençionamos nomear com três letras minúsculas para futura identificação no corpus:

- Entradas = **ent**
- Sopas = **sop**
- Pratos Principais
 

{	carne = <b>ppc</b>
	aves = <b>ppa</b>
	peixes e frutos do mar = <b>ppp</b>
	massas = <b>ppm</b>
- Acompanhamentos = **aco**
- Saladas = **sal**
- Sobremesas = **sbr**
- Tortas e Bolos = **teb**
- Pães = **pao**

A estrutura final de nosso corpus pode ser vista na **Figura 2**.

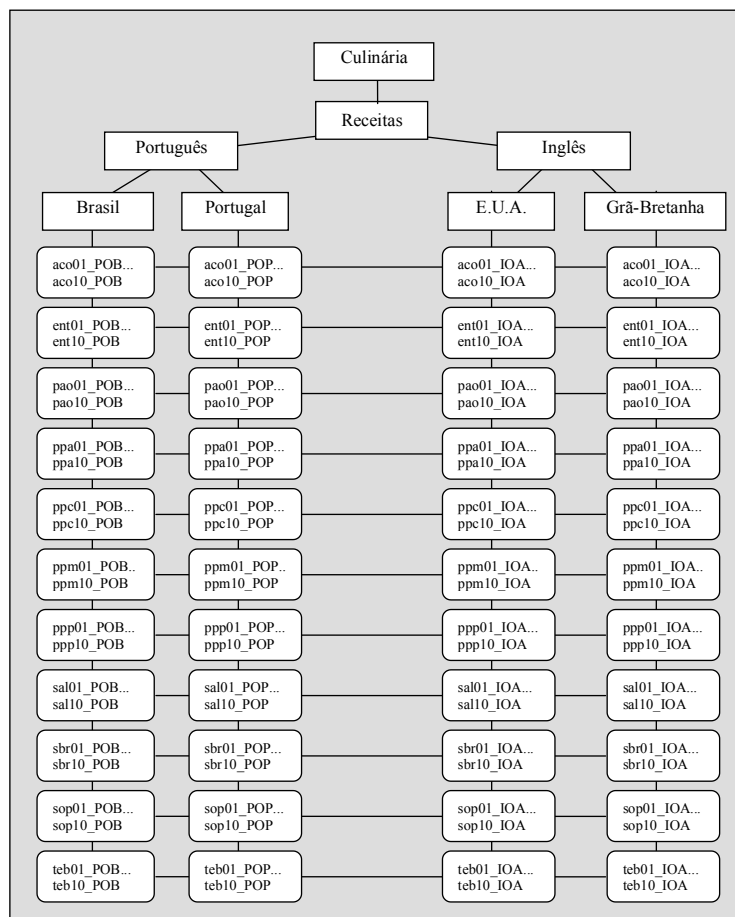
---

mel, lã, seda e couro. (<http://www.facom.ufba.br/prod/incomum/vegan.html>)

## 2.2 Suporte técnico

### 2.2.1 Hardware

Nosso corpus, por ser um projeto-piloto e por ser de área técnica, não é de grandes proporções, portanto não tivemos de nos preocupar com a aquisição de um computador específico – a primeira versão do corpus tem 73.372 bytes. Um ponto de grande importância para nós foi ter à disposição a Internet de banda larga, já que a coleta de dados e as pesquisas foram realizadas integralmente on-line.



**Figura 2:** Estrutura do corpus multivarietal de culinária português/inglês

### 2.2.2 Software

Para as buscas na Internet, utilizamos com maior frequência as ferramentas de busca Google<sup>12</sup> e WebCorp.<sup>13</sup> Para armazenar os textos, o Bloco de Notas do Windows, que grava os textos automaticamente em .txt, tipo de arquivo requerido pelo Wordsmith Tools<sup>14</sup> o qual, por sua vez, foi nossa ferramenta de análise do corpus.

Para a fase de organização dos dados obtidos, em tabelas, gráficos, esquemas, etc., usamos o Excel e o Word for Windows, da Microsoft.

## 2.3 Coleta e armazenamento do corpus

Os textos foram coletados na Internet com o recurso Copy/Paste e alocados diretamente num arquivo .txt já nomeado (p. ex. aco01\_POB) do Bloco de Notas. Finda essa etapa, o texto era editado, isto é, dele retirávamos manualmente tudo o que não fosse estritamente parte da receita (por exemplo, endereço eletrônico, links, créditos de fotos, espaços em branco). Em seguida o texto recebia o cabeçalho, as etiquetas (vide a seguir) e era salvo em sua respectiva pasta (POB, POP, IOA ou IOB).

### 2.3.1 Cabeçalho

O cabeçalho, cujo conjunto de informações categoriza cada receita e pode servir para delimitar pesquisas, inclui os seguintes campos:

- **<title>** título da receita **</title>**
- **<filename>** nome do arquivo, ex: aco01\_POB **</filename>**
- **<subcorpus>** “receitas” (futuramente podemos acrescentar “glossário” ou “artigo”, por exemplo) **</subcorpus>**

<sup>12</sup> [www.google.com](http://www.google.com)

<sup>13</sup> [www.webcorp.org.uk](http://www.webcorp.org.uk)

<sup>14</sup> Scott, M. (1999) *Wordsmith Tools* version 3. Oxford: Oxford University Press (versão demo disponível no site <http://www.lexically.net/downloads/download.htm>).

- **<collection>** “acompanhamentos”, “entradas”, “pães”, “pratos principais – aves”, “pratos principais – carnes”, “pratos principais – massas”, “pratos principais – peixes e frutos do mar”, “saladas”, “sobremesas”, “sopas” ou “tortas e bolos” **</collection>**
- **<language>** “POB”, “POP”, “IOA” ou “IOB” **</language>**
- **<mode>** “Internet” (futuramente podemos acrescentar “revista”, “livro”, etc.) **</mode>**
- **<status>** “caseira” (futuramente podemos acrescentar “profissional”) **</status>**
- **<publisher>** nome do *site* ou empresa que o mantém **</publisher>**
- **<editor>** nome do editor do *site* (pessoa ou empresa), quando identificado **</editor>**
- **<pubPlace>** endereço da receita na Internet **</pubPlace>**
- **<accessDate>** data de coleta da receita **</accessDate>**
- **<comments>** área reservada para comentários nossos **</comments>**
- **<name>** nome completo do autor da receita, quando mencionado **</name>**

A seguir, um exemplo de cabeçalho preenchido:

```

<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.shtml </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>

```

### 2.3.2 Etiquetas

Previendo estudos futuros, ou mesmo uma análise mais aprofundada das características textuais do corpus, decidimos acrescentar às receitas etiquetas “pragmáticas”, ou “discursivas”, para identificar as principais partes das receitas. Desse modo, poderíamos verificar, por exemplo, o número de ocorrências de um determinado vocábulo, por exemplo, “bolo”, excluindo sua ocorrência nos títulos das receitas. Também poderíamos contabilizar as ocorrências de verbos que indicam processos apenas quando juntos à lista de ingredientes, tais como “fatia\*” (fatiado, fatiada, em fatias, etc.).

Criamos, assim, as seguintes etiquetas:

- **<titRec>** título da receita **<titRec>**
- **<coment>** comentários do autor da receita, como “ótima para os dias quentes de verão” **<coment>**
- **<ingr>** lista de ingredientes **<ingr>**
- **<modFaz>** descrição dos procedimentos da receita **<modFaz>**
- **<class>** informações dadas pela receita quanto ao rendimento, tipo de prato, valor nutricional, tempo de preparo, custo, etc. **<class>**

A seguir, um exemplo de receita etiquetada:

```

<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 4:** Exemplo de texto etiquetado

A certa altura da construção do corpus encontramos uma receita especificando quais os tipos de equipamentos necessários para a sua execução, o que ocorreu, no total, 3 ou 4 vezes. Como não havíamos previsto essa categoria nas etiquetas, e já que sua ocorrência não foi significativa de modo geral, optamos por incluir tais informações no parâmetro “modo de fazer” (<modFaz>).

### 2.3.3 Dificuldades encontradas

A primeira dificuldade encontrada na coleta das receitas foi a existência de uma quantidade desproporcional de *sites* nas quatro variantes das línguas aqui consideradas. O número de *sites* disponíveis em inglês americano supera, de longe, os encontrados em português do Brasil, inglês britânico e português de Portugal, nessa ordem. Por sua vez, os *sites* de receitas de Portugal, além de serem em menor número, apresentam muitas receitas “regionais” e, com frequência, receitas “discursivas” (em que não há separação entre os “ingredientes” e o “modo de fazer” – um dos critérios seletivos que empregamos).

Ora, por um lado, essa questão é um fator a se considerar, já que pretendemos, justamente, destacar as diferenças e semelhanças entre as duas variantes do português. Por outro lado, para preservar a comparabilidade que desejávamos alcançar em nosso corpus, a coleta dessas receitas foi indesejável, pois, muito provavelmente, causaria um desequilíbrio na proporção “ingredientes/processos” na análise final do corpus. Além disso, o sistema de etiquetagem proposto teria de ser descartado. É certo, enfim, que tais questões merecem um estudo futuro.

Outra dificuldade foi distinguir as receitas britânicas das americanas, uma vez que muitos *sites* de receitas são mantidos e/ou alimentados por internautas, que muitas vezes não se identificam ao enviar sua colaboração. O uso de vocabulário distintivo (como “aubergine”, “courgette” – IOB, ou “eggplant”, “zucchini” – IOA) foi empregado, sempre que possível, para fazer essa identificação. Outra dificuldade para a identificação da origem da receita provinha da ausência da extensão, geralmente expressa no endereço do *site* (como “.br” para *sites* brasileiros, “.pt” para portugueses, “.uk” para britânicos e “Ø” para americanos). Há muitos *sites*, como os “.org” e “.com”, que não possuem essa terminação identificadora, o que significa dizer que estão excluídos de uma pesquisa em que o domínio era predeterminado na página de busca.

Encontramos também problemas de outras ordens. Um deles é o fato de algumas receitas estarem em formato PDF (i.e., podem ser lidas apenas com o programa Acrobat Reader), que demoram muito para abrir e, sendo muitas vezes protegidas por

senhas, tem-se impedida a cópia, embora seja permitida a sua impressão. Também se destaca o fato de que alguns *sites* mais elaborados usavam símbolos/figuras no corpo da receita (exs.: “\*\*\*”, para indicar o grau de dificuldade ou custo da receita; o desenho de uma panela para indicar o “modo de fazer”). Isso se revelou uma dificuldade a mais na etapa da cópia dos textos, pois em .txt todos esses símbolos se perdem e, por isso, têm de ser digitados manualmente – já que fazem parte da classificação da receita –, aumentando o tempo gasto na coleta.

Inicialmente tínhamos optado por fazer uma coleta aleatória de todas as receitas do português brasileiro numa dada categoria (tal como “pães”), depois iniciar a busca no europeu, em seguida no inglês americano e, finalmente, no britânico. No entanto, percebemos que, ao mudar de língua/variante, muitas vezes perdíamos tempo valioso procurando um equivalente para uma dada receita, sem resultado. Isso obrigava-nos a descartar a receita já encontrada, trabalhada e etiquetada e substituí-la por outra, que pudesse ser encontrada nas outras variantes. Decidimos, então, promover a seleção de todas as receitas em português e, em seguida, todas em inglês. Mas, ainda assim, o problema persistia ao mudarmos de língua. Por isso, depois de muito esforço e paciência para completar dois ou três subcorpora, decidimos fazer um planejamento prévio das receitas que seriam incluídas, com base numa pesquisa rápida nos principais *sites*, evitando, assim, incluir receitas para as quais não seria possível (ou fácil) encontrar equivalentes nas quatro variantes.

Durante a coleta dos textos anotamos todas as dificuldades encontradas, listando o vocabulário desconhecido, o qual, inclusive, poderá servir de base para uma pesquisa futura. Todo o trabalho foi realizado entre 02/12/2002 e 07/02/2003. A seguir, faremos alguns comentários sobre as dificuldades específicas de cada subcorpus, por ordem alfabética das siglas empregadas (Vide ANEXO I para uma listagem das receitas coletadas).

### 2.3.3.1 Acompanhamentos

Este subcorpus apresentou grande dificuldade, já que os acompanhamentos abrangem um leque muito amplo de recei-

tas. O problema maior ocorreu por conta das não-equivalências culturais, por exemplo, a presença marcante de “batatas” na cultura de língua inglesa, ou a presença do arroz/feijão na cultura brasileira. Alguns acompanhamentos clássicos numa cultura não existiam na outra, como é o caso da “farofa” brasileira, das “migas” portuguesas ou dos “baked beans” americanos.

Aparentemente, o correspondente mais próximo da farofa em Portugal são as “migas”, embora muitas vezes a receita dê mais a impressão de um “cuscuz paulista”, ou um pirão espesso, só que engrossado com “pão ralado” (farinha de rosca). Como são servidas não somente com peixe, como é o caso mais comum do pirão no Brasil (as migas podem, inclusive, constituir uma refeição completa), resolvemos criar o paralelismo das “migas” com a “farofa” brasileira. No caso do inglês, as receitas de “Brazilian style ‘farofa’” são em grande número, mas por não sabermos se se tratavam de tradução ou não, escolhemos o cuscuz marroquino para seu equivalente. Apesar de o cuscuz ser um prato ‘estrangeiro’, já está tão ambientado na cultura dos países de língua inglesa que seria difícil deixá-lo de fora da categoria “acompanhamentos”.

Outros pratos que julgamos fundamental incluir, como mencionado acima, foram o *arroz* e o *feijão* brasileiros. Por sorte encontramos os dois numa única receita, “arroz com feijão”, em um *site* de ‘Dicas para noivas’. Em Portugal foi fácil encontrar prato semelhante, já que há muitas receitas de “Arroz de feijão”, em que os dois são cozidos juntos. Mas no inglês foi complicado. Devido à influência mexicana, a busca foi mais rápida no inglês americano, mas no britânico, depois de longo tempo de pesquisa, tivemos de nos contentar com um “Mexican rice and bean casserole”, encontrado no *site* da BBC. Isso parece indicar que nem a cominação de *arroz* e *feijão* nem a versão mexicana do prato são comuns na cultura britânica.

### 2.3.3.2 Entradas

Inicialmente havíamos pensado em criar subcorpora em duas categorias: “Petiscos” e “Entradas”, por acreditarmos que os pratos que apareceriam em cada uma delas seriam muito

distintos. No entanto, durante a pesquisa prévia para a elaboração do roteiro de receitas a serem coletadas, percebemos que há certa confusão em relação aos pratos dessas categorias. A situação é mais complexa no inglês – um *site* britânico distingue “hors d’oeuvre” de “party food” e de “appetizer”. Mas ao verificarmos as receitas, vimos que muitas delas se repetem nas diferentes categorias – “codfish balls”, por exemplo, é “hors d’oeuvre”, “party food” ou “appetizer”? Nossas pesquisas revelaram que ‘vai do gosto do freguês’. Por isso, resolvemos incluir todos esses “petiscos, acepipes, abrideiras, *appetizers*, *starters*, *hors d’oeuvre*, *party food*”, etc. na categoria “entradas” – entenda-se, com isso, ‘tudo o que é salgado, não constitui refeição completa e não é nem prato principal, nem sopa, nem salada’.

Como mencionado acima, as correspondências nessa categoria não foram fáceis de se encontrar, o que revelou uma grande diferença cultural entre os países de língua portuguesa (em que há muitas e variadas receitas do gênero, inclusive os “salgadinhos”, inexistentes como tais no inglês) e de língua inglesa. No caso dos “croquetes”, por exemplo, observamos no inglês americano uma alta ocorrência de “Salmon croquettes”, o que nos fez optar por essa receita em detrimento de uma ou duas ocorrências de “Meat croquette” (as numerosas “meat balls” correspondem às nossas “almôndegas” – um tipo de prato principal).

Outro caso interessante foi o do “bolinho de bacalhau”. Após longo tempo de busca na Internet, e inconformados de não encontrar uma única receita do famoso prato português, nos demos conta de que os “bolinhos”, em Portugal, são conhecidos também como “pastéis”. Por fim, encontramos os “Pastéis de bacalhau à Portuguesa”. Já no inglês proliferavam as receitas de internautas portuguesas colaborando com sua versão particular da receita, mas preferimos não coletá-las por não sabermos se eram traduções ou não de receitas portuguesas (uma delas, disponível em vários *sites*, era justamente a tradução literal daquela apresentada no *site* português de receitas e que havíamos coletado. Essas receitas, no entanto, podem ser de interesse futuro para constituir um corpus de receitas traduzidas).

### 2.3.3.3 Pães

Essa foi, sem sombra de dúvida, a categoria mais difícil. Isso se deve a vários motivos: praticamente não há na Internet receitas de pães em português de Portugal (notamos que o que chamam de “broa” e “bolo” muitas vezes é algo parecido com nosso “pão”, mas metade das poucas receitas encontradas: a) é do tipo “discursivo”, ou seja, não separam os “ingredientes” do “modo de fazer”; e b) são muito regionais, isto é, constam em sites de comidas típicas de certas regiões). Quando digitávamos “pão pães” como palavra de busca, e “.pt” como restrição de domínio, o programa de busca retornava milhares de receitas. No entanto, em 99% delas o pão era ingrediente (*pão ralado, pão migado*, etc.) e não a receita do pão em si.

Quanto aos americanos (e britânicos também), com o advento das *breadmachines* proliferaram receitas de pães que se limitam a uma lista de ingredientes a serem colocados na máquina de fazer pão, e pronto! Não que tenha sido difícil achar receitas de pão ‘tradicionais’ em inglês, o difícil foi distinguí-las das outras no momento da busca. Isso nos tomou bastante tempo.

Um terceiro motivo para a grande dificuldade com a categoria pães é novamente de cunho cultural. Ao lermos as receitas de “muffin”, por exemplo, tínhamos a certeza de estarmos diante de um “mini-bolo”. Os americanos e ingleses, no entanto, têm outra opinião: para eles os “muffins” são um tipo de “pão” e, de acordo com nossa categorização de receitas, entrariam aqui, e não em “bolos e tortas”, como mandaria nosso julgamento. Tomamos uma atitude não muito científica, correndo o risco de comprometer a representatividade de nosso corpus (risco que assumimos por tratar-se de um projeto piloto): ignoramos os *muffins*, assim como as *pizzas*, por não termos chegado à conclusão ainda sobre qual categoria eles pertencem. O clássico “Sourdough bread” também ficou de fora, por não encontrarmos equivalente em português do Brasil. Enfim, de modo geral, em nossas buscas procuramos nos guiar mais pelos ingredientes e processos, e não tanto pelos nomes.

Uma exceção foi a “broa” – a portuguesa parece ser um tipo de pão salgado feito com “farinha de milho” (que não sabe-

mos ainda se é o nosso “fubá”), ao passo que a nossa é, em muitas regiões do Brasil, considerada uma quitanda doce. O correspondente que encontramos em IOA, apesar de levar *cornmeal* e ter *bread* no nome, leva fermento químico, é de massa mole, mais parecido com um bolo. Já em IOB, a receita leva fermento biológico e fubá, mas leva também “8 teasp molasses”, o que talvez o torne um pouco doce.

Não foi fácil, também, encontrar uma receita de pão com fermento químico em POP, basta ver pelo nome da receita que por fim selecionamos, “Pão pequenino”. Isso dá uma idéia do quanto as pistas eram difíceis de rastrear. O representante do IOA escolhido, “Irish soda bread”, a princípio não nos parecia correto – se estamos falando de receitas americanas, como pode haver um “Irish” em seu título? – mas depois concluímos que, assim como o nosso bolo inglês, a ‘nacionalidade’ do pão refere-se muito mais aos ingredientes e a um processo específico de execução da receita (no caso do “bolo inglês”, à forma em que é assado) do que propriamente ao país/cultura de origem. Assim, decidimos considerá-los equivalentes.

Por fim, gostaríamos de registrar nosso espanto: parece inacreditável, mas não encontramos uma receita sequer de *pão francês* ou *baguette* em POP, embora eles sejam citados em várias receitas como ingrediente, tal como mencionamos acima. Tentamos todas as palavras de busca que pudemos imaginar para chegar a essas receitas, mas não obtivemos resultado. Por fim, decidimos colocar como representante uma receita qualquer da categoria em POP – escolhemos o “Pão de aveia e mel” – muito a contragosto...

#### 2.3.3.4 Pratos Principais

##### a) Aves

O maior problema encontrado nessa categoria foi a diversidade de nomes dados aos diferentes tipos de frango, que podem variar conforme a idade, a alimentação, a forma de confinamento, o peso, se ele é castrado ou não, entre outros (assunto que merece um estudo à parte). Por ora achamos importante comentar que não conseguimos identificar o correspondente de



“galeto” ou “frango de leite” no POP e, portanto, a receita que foi colocada lado a lado com a de “galeto” brasileira foi selecionada pela menção de “1 frango pequeno”. Ao que tudo indica, há, no mínimo, dois correspondentes para o “galeto/frango de leite” em inglês: (a) “(rock) cornish (game) hens” (“poussin” em francês, presente apenas em receitas de chefs), que são semelhantes no tamanho e na textura da carne às (b) “spring chickens”, (que apareceram também como “fryer”, “broiler-fryer chicken”, ou ainda “baby chicken”). Os primeiros resultam de um cruzamento de duas raças de frango que, mesmo adultos, são pequenos – seriam semelhantes ao nosso “garnisé”? O dicionário Houaiss parece confirmar a suspeita:

*“garnisé. adjetivo e substantivo de dois gêneros. 1 Rubrica: zootecnia. diz-se de ou galináceo de porte muito pequeno, pertencente a diversas raças, cujos primeiros casais trazidos para o Brasil vieram da ilha Guernsey, na Grã-Bretanha”*

Mas essa equivalência não ocorre com as receitas, ou seja, não foi encontrada nenhuma com essa variedade de ave em POB ou POP. No caso do “frango / galinha caipira” (“free-range chicken”), o mais próximo que achamos em POP foi “galinha caseira”.

Não conseguimos encontrar, também, o correspondente de “chester” em inglês (!). Já o frango defumado não figura em nenhuma receita de “prato principal” em POP, IOA ou IOB, por isso foi descartado de nosso planejamento inicial. Interessante foi a constatação de que nas quatro variantes havia várias receitas de “capão” – frango capado e alimentado com dieta especial, muito usado nas festas de Natal antes da popularidade do peru. Assim, acrescentamos a receita ao corpus.

Outra ave problemática foi a codorna, pois parece haver dois ou três tipos diferentes, sendo que o tipo mais comum no Brasil (*Coturnix coturnix*) é diferente do que é comum entre os americanos e britânicos (*Nothura maculosa*, que eles chamam de “quail”). O Aurélio dá “codorniz” como sinônimo de codorna, mas cita outras duas aves: codorna-buraqueira e codorna-mineira – nenhuma das duas com o nome científico igual ao de “quail”. Já o Dicionário Houaiss deixa o problema ainda mais

complicado. Ele cita, além da “codorna”: “codorna-amarela, codorna-buraqueira, codorna-caseira, codorna-comum, codorna-doméstica, codorna-pintada, codorna-mineira, codorniz e codornizão”. No verbete de **codorniz** lê-se:

Rubrica: ornitologia.

- 1 m.q. **codorna** (*Coturnix japonica*)
- 2 m.q. **codorna-comum** (*Nothura maculosa*)
- 3 m.q. **codorna-mineira** (*Nothura minor*)

Complicações à parte, o que nos importa no momento é que, em se tratando de receitas, no Brasil, conforme atesta sua frequência na Internet, usa-se mais “codorna” que “codorniz” e, em inglês, *quail* é mais usado do que *bobwhite* (dado como sinônimo pelo Dicionário Webster). Com base nessas constatações culturais refletidas no uso dos termos, estabelecemos a equivalência entre as receitas, mas essa é uma questão que, por merecer maior atenção, será objeto de outro estudo.

#### b) Carnes

A não-equivalência entre os cortes de carne nas quatro variantes, e mais especificamente no Brasil em relação aos outros países, foi o motivo de maior dificuldade no tratamento dessa categoria. O “cabrito” não foi incluído, porque não encontramos nenhuma receita com o equivalente dicionarizado em inglês “kid / baby goat”. Optamos por substituí-lo pela vitela (“veal”), que aparece em muitas receitas na rede.

#### c) Massas

Os nomes italianos para as massas prevalecem nas quatro línguas, embora a grafia muitas vezes sofra alterações de uma língua para a outra, ou mesmo dentro da mesma língua: “lasanha / lasanha / lasagna” (português), “lasagna / lasagne” (inglês); “cap(p)el(l)et(t)i” e “fet(t)uc(c)ini/e” (ambas as línguas), etc. Devido à sua grande variedade, procuramos nos ater aos ingredientes e ao modo de fazer das receitas.

As massas recheadas não parecem ser muito populares em Portugal. Para o “cap(p)el(l)et(t)i” encontramos apenas duas ou três receitas na rede, em geral mal escritas. Um equivalente

possível é o “tortelloni”, mas em receitas só foi encontrada uma ocorrência com “tortellini”. No IOB havia uma receita de “tortelloni” cuja foto exibia o que chamamos de “cap(p)el(l)et(t)i” no Brasil (equivalência confirmada no corpo da receita).

Ressaltamos, novamente, que a “pizza” não foi incluída aqui, já que não nos decidimos quanto a considerá-la *prato principal*.<sup>15</sup>

#### **d) Peixes e frutos do mar**

A quantidade de nomes de peixes parece quase infinita, razão pela qual optamos por nos ater aos mais populares, dando atenção especial ao modo de fazer. Os moluscos de concha e os diversos tipos de camarão são outro assunto, que merece ser estudado a fundo, posto que não foi possível fazê-lo no presente trabalho. Afora isso, os peixes e frutos do mar não apresentaram grande dificuldade para o estabelecimento das equivalências.

##### *2.3.3.5 Saladas*

Os textos envolvendo as saladas não apresentaram grande dificuldade. Preocupamo-nos, especialmente, em garantir ampla variedade de ingredientes, já que o processo não é um ponto central nesse tipo de receita.

##### *2.3.3.6 Sobremesas*

Essa foi outra categoria que apresentou grande dificuldade. O “manjar branco” (sbr01), o “pudim de pão” (sbr02), o “pudim de leite” (sbr09) e o “arroz doce” (sbr03), em inglês, acabaram classificados numa mesma subcategoria – a dos “custards” e “puddings”. Pelo que pudemos perceber, “custard” tem signi-

<sup>15</sup> O conceito de *prato principal* pode não ser consensual nas quatro variantes. Neste trabalho consideramos *prato principal* aquele que é o mais substancial da refeição e que, em geral, é servido com acompanhamentos como legumes, arroz, batatas e/ou saladas.

ficados distintos em IOB (molho/creme de baunilha) e IOA (pudim).

O *bombocado*, por exemplo, foi escolhido por ter um cognato – “Bom bocado” – em POP: as receitas mostram, no entanto, que são sobremesas totalmente distintas. Não conseguimos para o prato um equivalente em inglês; os que mais se aproximavam dele eram alguns tipos de torta, além dos *cheesecakes* (que entrariam na categoria “tortas e bolos”) e, novamente, os *custards*. Optamos por uma receita de “Coconut cream cheese flan”, em IOA, e “Peggy’s pudding”, que leva coco na composição, para IOB.

Quanto ao doce feito de abóbora, ao digitarmos, no inglês, as palavras de busca “pumpkin squash sugar”, recebíamos uma infinidade de receitas em inglês, das quais 99,9% eram de “Pumpkin pie”, da categoria das tortas. O *pavê* também foi difícil de se encontrar: no português europeu foi pela escassez de receitas de doces, mas em inglês não tínhamos nenhuma pista de como encontrar essas receitas, ou como realizar as buscas. Começamos pelas palavras de busca “cracker sugar cream”, mas as muitas receitas exibidas pediam a bolacha/biscoito “esfarelada/o” (*crushed*), processo muito usado para fazer base de tortas (mais uma vez as “tortas e bolos” atravessavam nosso caminho!). Depois de várias tentativas, a busca com “layer sugar cream cracker crushed” começou a oferecer melhores resultados e, assim, encontramos “Seven-layer icebox cake” em IOA e “Brandy mocha trifle” em IOB.

Curiosamente, não conseguimos uma única receita de pudim de leite em POB que não levasse “leite condensado”; por isso optamos pelo “Pudim de chocolate”, que é feito de “leite” e “ovos”, como nas receitas equivalentes encontradas para as outras variantes. Em IOA o processo também foi demorado, pois a busca com “egg custard” praticamente não retornou resultados. Acabamos encontrando uma receita com esse nome por meio da busca pelos ingredientes em um *site* cujos títulos das receitas provavelmente não estavam indexados, de forma que as ferramentas de busca não conseguiam localizá-los com a palavra de busca “egg custard”, embora se tratasse justamente do título dado à receita.

### 2.3.3.7 Sopas

As sopas foram a primeira categoria que coletamos e somente por esse motivo houve um alto índice de perdas (receitas sem correspondentes, que foram trocadas por outras), já que ainda não tínhamos começado o planejamento prévio de coleta. No geral, não apresentou grandes dificuldades, mas a “canja de galinha” merece um comentário: não conseguimos encontrar qualquer receita de sopa que levasse arroz, frango e legumes em IOB. Optamos pelo paralelismo com o ingrediente “chicken”, em detrimento de “rice”.

### 2.3.3.8 Tortas e Bolos

Ao coletarmos as receitas dessa categoria ficamos na dúvida se a incorporávamos às “Sobremesas” ou não, já que muitas tortas, e até mesmo bolos (especialmente nos Estados Unidos, até onde sabemos), são muitas vezes servidos como sobremesa. Mas dada a complexidade encontrada na coleta das receitas da categoria “Sobremesas”, decidimos manter as duas separadamente, também para que os “doces” ficassem mais bem representados no corpus (as outras 9 categorias são essencialmente de pratos salgados).

Destacamos aqui o caso da torta doce de queijo. Seria fácil encontrar uma receita do tradicional “cheese cake” do IOA nas outras variantes (o que de fato chegou a ser feito no POP e no IOB), mas a escolha nos pareceu um tanto falsa para o POB, já que no Brasil a iguaria não chega a fazer parte da nossa cultura (ainda?). A opção foi, portanto, manter a equivalência somente no ingrediente genérico “queijo”. Com isso encontramos em POB a conhecida “Torta de ricota”, uma “Queijada serrana” em POP (receita bastante comum nos *sites* POP) e a “Yorkshire curd tart”, tradicional no IOB.

A coleta de textos por categoria deixou bem clara a necessidade de se fazer um levantamento das receitas mais comuns em todas as categorias para cada variante. Somente assim conseguiríamos um corpus comparável no nível cultural, pois o que

ora foi compilado acabou por focar mais os aspectos léxicos e sintáticos, justamente os aspectos em que nossos questionamentos iniciais se concentravam.

## 2.4 Processamento do corpus

Dentre as fases da elaboração de um corpus propostas por Atkins *et al* (1992), “corpus processing” refere-se à etapa de preparação do corpus para que possam ser extraídas dele as informações de interesse. É nessa fase que entram em cena as ferramentas de processamento de corpora. Os autores mencionam ainda, em seu texto, as ferramentas mais sofisticadas para a tarefa, tais como os etiquetadores gramaticais e sintáticos, lematizadores, etc. Para os propósitos deste trabalho foi usado apenas o programa de análise lingüística *Wordsmith Tools* para avaliar o corpus, conforme comentaremos a seguir.

### 2.4.1 Types, Tokens e Type/Token Ratio

*Token* é cada palavra do corpus, reconhecida pelo programa como aquilo que figura entre espaços em branco. Funciona como o “Contar palavras” do *Word for Windows*, sendo que o único critério empregado para contabilizá-las é o recurso físico mencionado acima (os espaços). Já *Type* é cada palavra distinta que pode se repetir 2, 10 ou 100 vezes até perfazer o total de *Tokens*. Assim, na frase “a banana nanica é mais saborosa que a banana maçã” temos 10 *Tokens* (palavras) e 8 *Types* (palavras distintas), pois “banana” e “a” aparecem duas vezes no exemplo.

O *Type/Token Ratio* é um cálculo utilizado para determinar a riqueza lexical de um corpus. O *Wordsmith Tools* calcula essa razão com uma regra de 3:

$$x = \frac{\text{no. de Types} \times 100}{\text{no. de Tokens}}$$

Por esse cálculo, quanto maior o *Type/Token ratio*, maior a riqueza lexical.<sup>16</sup> No exemplo acima a razão seria de 80% (8 x 100 ÷ 10), ou seja, para cada 100 palavras, 80 são distintas e 20 repetidas. Se o exemplo fosse “a banana nanica e a banana maçã são inferiores à banana ouro e à banana prata em sabor e nutrientes” teríamos:

<i>Tokens</i> : 20	<i>Types</i> : 12	<i>Type/Token ratio</i> : 60%
--------------------	-------------------	-------------------------------

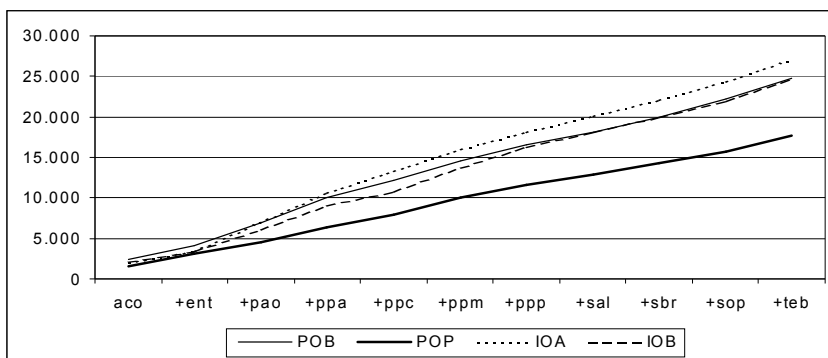
Elaboramos uma planilha em Excel com os valores de *Type*, *Token* e *Type/Token Ratio* para cada receita do *site*, cada subcorpus e para o corpus total. Os dados foram extraídos do corpus com a ferramenta “*Wordlist*” do *Wordsmith Tools*, que apresenta três listas diferentes: uma com os *Types* em ordem decrescente de frequência, outra com os mesmos *Types* em ordem alfabética, e uma terceira, com a estatística dos arquivos examinados, de onde foram retirados os dados para a planilha.

Uma breve análise dos valores da tabela permite-nos fazer diversas constatações:

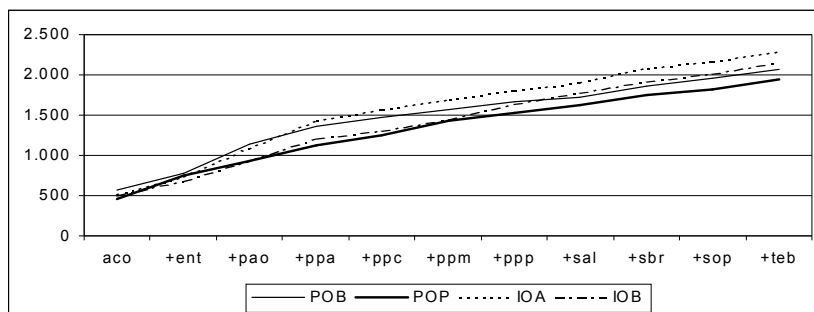
- IOA é o corpus que tem maior número de palavras (*Tokens*), seguido pelo POB, IOB e, bem abaixo da média dos 25.000, pelo POP. No entanto, o mesmo não ocorre com os *Types*: POP apresenta o maior índice de riqueza lexical (11%), seguido por IOB, IOA e POB. Mas essa inversão não é casual. Conforme demonstra Biber (1993), o aumento do *Type/Token ratio* é inversamente proporcional ao aumento dos *Types*; isto é, quanto mais palavras no corpus, maior a probabilidade de repetição. Isso está mais bem representado nos Gráficos 1, 2 e 3 abaixo, em que o eixo “x” representa cada adição de uma nova categoria dos subcorpora ao corpus total, e o eixo “y” representa os valores para, respectivamente, *Type*, *Token* e *Type/Token ratio* a cada nova adição. Observa-se que, en-

<sup>16</sup> Outra forma de calcular o *Type/Token Ratio* é dividir o número de *Tokens* pelo número de *Types*. Nesse caso, quanto maior o número obtido, menor a riqueza lexical. Por isso, é importante explicitar a fórmula utilizada para que se possa interpretar os resultados corretamente.

quanto os *Tokens* (**Gráfico 1**) têm uma trajetória constante e regular (com o acréscimo de cada categoria o corpus aumentava em cerca de 2.000 palavras), os *Types* (**Gráfico 2**) mostram uma ascensão brusca até **ppa** (momento em que os subcorpora somavam aproximadamente 9.000 palavras) e depois dessa categoria, um crescimento mais constante, que vai diminuindo conforme aumenta o número de *Tokens* e as palavras começam a se repetir mais vezes dentro de cada subcorpus. O **Gráfico 3**, que representa o *Type/Token ratio*, apresenta uma trajetória descendente, pelos motivos que salientamos acima. Por sua vez, o POP, com o menor número de palavras representadas no corpus, destaca-se dos outros, mantendo valores mais altos em todos os estágios da construção do corpus, justamente por ter menos *Tokens*.

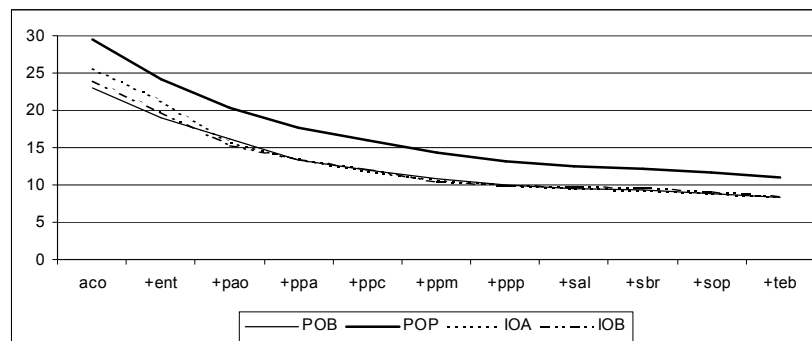


**Gráfico 1:** Tokens (evolução)



**Gráfico 2:** Types (evolução)



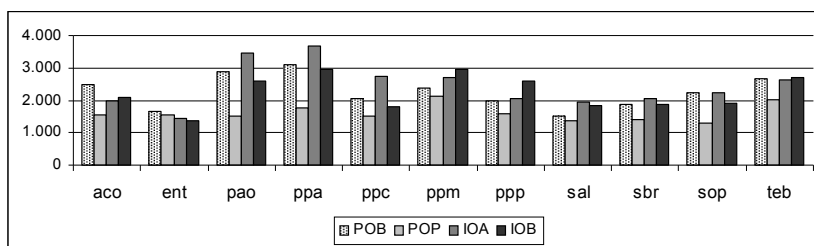


**Gráfico 3:** Type/Token Ratio (involução)

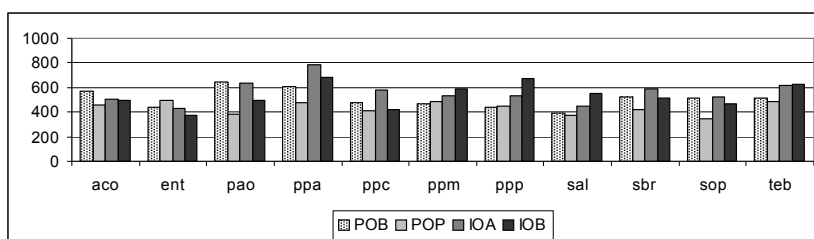
- Há certas receitas e categorias de receitas que apresentam um número de *Tokens* claramente diferenciado. É o caso, por exemplo, da receita **ent02**. Uma consulta ao ANEXO I, nesse caso, desfaz o mistério: na verdade, trata-se da receita do “Patê de atum”, que não apresenta nem um grande número de ingredientes nem complexidade no preparo, daí os baixos valores. Já a categoria “Pães” é uma das que têm maior número de *Tokens* (com exceção de POP, para o qual, como mencionamos anteriormente, foi muito difícil encontrar receitas nessa categoria), e isso se justifica, provavelmente, pela complexidade do preparo de pães.

Todos esses aspectos quanto às diferenças entre as diversas categorias podem ser observados, em detalhe, nos Gráficos 4, 5 e 6 abaixo. O **Gráfico 4** mostra os valores representativos do número de *Tokens* em cada subcorpus. Por ele é possível observar a tendência das quatro variantes em apresentar receitas com poucos *Tokens* na categoria “Saladas”, possivelmente pelos mesmos motivos apresentados acima para **ent02**. Observa-se também que, enquanto POP e IOB apresentam um maior valor de *Tokens* para **ppm**, se comparados a outras categorias nas mesmas variantes, IOA e POB mantêm relativamente o mesmo número de *Tokens* de outras categorias. Seria esse um sinal de que a culinária italiana está bem mais enraizada nas culturas brasileira e americana do que na portuguesa e britânica, e daí a necessidade maior de ‘explicações’ nas receitas desses países? O **Gráfico 5** mos-

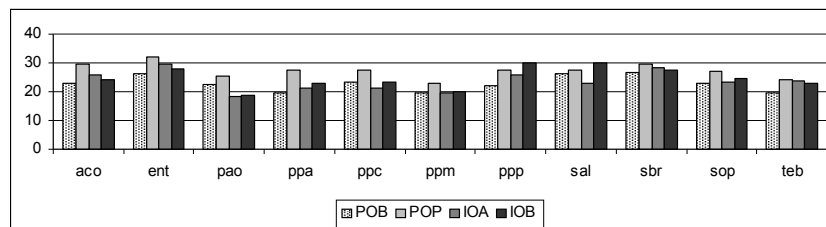
tra os *Types* por categoria. As colunas exibem, no geral, valores semelhantes aos apresentados no **Gráfico 4**, com algumas exceções interessantes: enquanto no gráfico de *Tokens* a categoria **ppp** é representada por uma ligeira diminuição em todas as variantes, neste tal categoria é representada por uma ascensão expressiva em IOB. É possível que, apesar de terem poucos *Tokens*, as receitas de “Peixes e frutos do mar” conttenham ingredientes muito variados, os quais nem sempre se repetem em outras categorias. Isso pode ser comprovado no **Gráfico 6**, em que a relação *Type/Token* está representada nas diferentes categorias. Há evidências de que, tanto em “Pratos principais (peixes e frutos do mar” (**ppp**) quanto em “Saladas” (**sal**) e “Entradas” (**ent**), a riqueza lexical é grande se comparada a categorias como **pao** ou **ppa**, que apresentaram altos valores de *Token*, mas porcentagens mais baixas para *Type/Token Ratio*.



**Gráfico 4:** Tokens (por categoria)

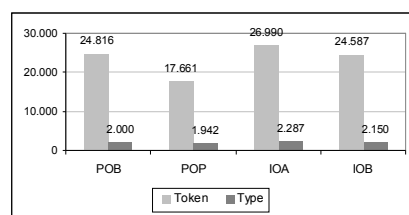


**Gráfico 5:** Types (por categoria)

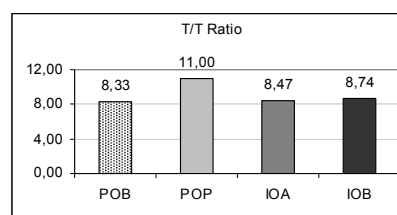


**Gráfico 6:** Type/Token Ratio por categoria

- Os **Gráficos 7 e 8** permitem que, mais uma vez, sejam confirmadas tanto a diferença de resultados obtidos em POP com relação às outras três variantes quanto a relação inversamente proporcional que há entre número de *Tokens* e riqueza lexical (*Type/Token Ratio*) nas receitas coletadas.



**Gráfico 7:** Tokens e Types (por variante)



**Gráfico 8:** Type/Token Ratio (por variante)

#### 2.4.2 Wordlists e Keywords

Conforme mencionado acima, “*Wordlist*” é uma ferramenta do *Wordsmith Tools* que faz uma listagem de todas as palavras dos arquivos selecionados, agrupando as palavras repetidas para calcular sua frequência conjunta nos textos. Outra ferramenta do mesmo programa é a “*Keyword*” que, grosso modo, seleciona, por comparação com uma segunda *Wordlist*, as palavras cuja frequência no corpus de estudo é significativamente maior ou menor do que no corpus de referência<sup>17</sup> – i.e., corpus usado para a comparação. Por exemplo, se comparamos a

<sup>17</sup> O ideal é que o corpus de comparação seja maior que o corpus de estudo.

*Wordlist* de um artigo de astronomia com a de um corpus de editoriais de jornal, a *Keyword* apresentará as palavras que só ocorrem, ou ocorrem com uma frequência muito maior, no artigo de astronomia (*Keywords* positivas) e aquelas que, porventura, ocorram com maior frequência ou exclusivamente no editorial (*Keywords* negativas). Isso pode ser usado, por exemplo, para se determinar o vocabulário específico de uma dada área, ou seja, termos que ocorrem especificamente naquele contexto ou que, justamente, não ocorrem naquele contexto.

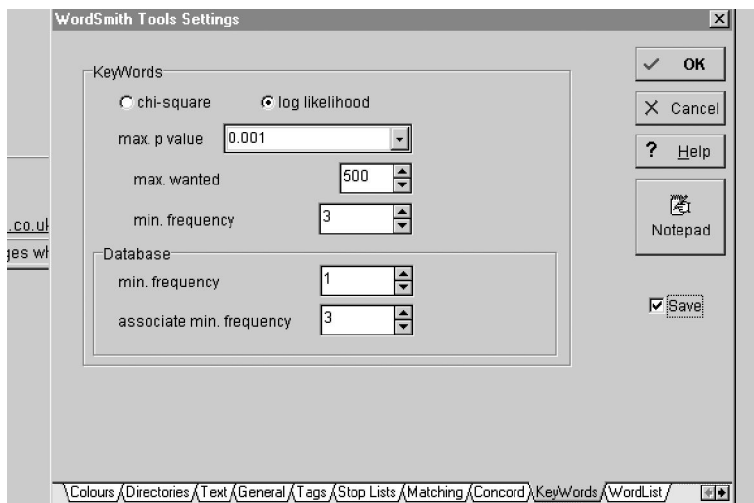
Para esse estudo, foi necessário elaborarmos as seguintes *Wordlists*:

- do total de cada uma das quatro variantes;
- das seções da receita, por exemplo “modo de fazer” (modfaz), “ingredientes” (ingr), etc. em cada uma das variantes;
- do total de cada variante menos a seção da receita a ser estudada,<sup>18</sup> por exemplo, POBtotal menos o “modo de fazer”, POBtotal menos os “ingredientes”, etc., o que resultou em 20 *Wordlists*, 5 em cada variante;
- de cada categoria em cada uma das línguas e variantes, o que resultou em 44 *Wordlists*;
- do total de cada variante menos a categoria a ser estudada, ou seja, POBtotal menos acoPOB, POBtotal menos entPOB, etc., totalizando mais 44 listas.

O passo seguinte foi comparar essas listas (recurso “*Compare 2 Wordlists*”, do *Wordlist*, que faz a *Keyword* dos arquivos escolhidos). Para tanto, ajustamos os critérios de pesquisa de keywords da seguinte forma:<sup>19</sup>

<sup>18</sup> Ao se extrair a lista de palavras-chave (*keywords*), os textos do corpus de estudo não devem fazer parte do corpus de comparação.

<sup>19</sup> O único valor alterado aqui (os outros são *default*) é o de P, que é uma medida estatística usada para calcular a probabilidade de o evento em questão ter ocorrido por acaso. Considerando-se que em Ciências Humanas o maior P aceitável é 0,05, ou seja, 5 ocorrências em 100, ou 1 em 20, aumentamos o valor de P para 0,001 – uma chance em mil de ser casual a presença daquele item lexical na lista de palavras-chave do corpus analisado. Dessa forma, conseguimos contemplar itens importantes os quais, caso aplicássemos um P menor, seriam desprezados.



Procedemos então às seguintes comparações:

**A.** cada categoria com o corpus total daquela variante, menos a categoria estudada. Ou seja, comparar, por exemplo **aco...POB**, **ent...POB**, **ppm...POB**, etc. com o **POBtotal** sem a categoria em questão, e assim sucessivamente com as 44 categorias, 11 de cada língua. Disso resultaram pequenas listas de *Keywords* em que figurava o vocabulário específico – se havia algum – daquela categoria de receitas em relação ao conjunto delas numa mesma variante (vide ANEXO II para uma amostra das categorias massas e sobremesas). Destacamos, a seguir, alguns dos resultados que mais nos chamaram a atenção:

- No caso das massas, alguns itens estão presentes em todas as variantes, como “queijo”, “lasanha”, “parmesão” e “ralado”. Curiosamente, “tomate(s)” não aparece(m) nas keywords de POP. Seria esse um indício de que é largamente empregado em outros tipos de receitas em Portugal? Ou seria um problema da quantidade menor de types e tokens das receitas POP, o que interferiria nas estatísticas?
- As palavras-chave mais presentes nas listas para POB e IOA se referem a diferentes tipos de queijo, massas, molhos, bem como ingredientes que podem ser usados

TRADTERM, **10**, 2004, p. 313-358

nesses molhos. Seria isso um indício de que, nessas culturas, o macarrão já está mais incorporado ao dia-a-dia, estimulando a criação de novas formas de apresentação e/ou modos de preparo? Seria isso um indício de uma maior influência da cultura dos imigrantes italianos nesses países?

- A presença de “flour” nas *Keywords* negativas de IOA pode ser uma indicação de que, nessas receitas, a massa é comprada pronta – o que pode indicar um fator cultural americano: a presença do “fast food”, que privilegia os alimentos semiprontos;
- A ausência de “sugar” nas *keywords* negativas do IOA pode ser justificada pela presença de “Brown sugar” (2 ocorrências) no molho da receita de *lasagna*, conforme verificamos no corpus, com o uso da ferramenta Concordance;
- No caso das sobremesas, é curioso observar a presença do “leite” + “condensado” nas *Keywords* positivas de POB e POP. O mesmo não ocorre, porém, com o correspondente nas variantes do inglês, o que pode ser uma indicação de que o ingrediente é mais difundido entre os países de língua portuguesa. O mesmo pode ser dito da palavra – que julgamos tratar-se do substantivo – “doce” nessas mesmas listas. Em contrapartida, Portugal parece ser o único a resistir ao “chocolate” (que nas receitas brasileiras é representado pelo “achocolatado”). Observa-se também que os países de língua inglesa possuem várias receitas de sobremesa em que o café e as bebidas alcoólicas são usados como ingredientes. No entanto, o tamanho reduzido do corpus não nos permite fazer tal afirmação sem ressalvas – a presença ou ausência de ingredientes nessas listas podem estar relacionadas à coleta dos dados. Por isso, somente num corpus mais representativo elas poderiam ser apontadas, com certeza, como indício de diferenças culturais;
- A metodologia empregada aqui permitiu-nos identificar vários equivalentes de tradução. Permitiu também identificarmos a grafia diferenciada de alguns termos,

como por exemplo, a de lasanha em inglês (IOA – “lasagna”, IOB – “lasagne”), ou do “brócolis” brasileiro, que em português de Portugal é “brócolos”. Essas distinções são de grande importância, tanto para a tradução quanto para a versão.

**B.** as diferentes partes da receita em cada variante com o corpus total, menos a seção da receita analisada. O objetivo era verificar se havia diferença no vocabulário específico para cada seção da receita e, em caso afirmativo, contrastar os resultados obtidos entre as variantes de cada língua, ou seja, POB x POP e IOA x IOB (vide ANEXO III para uma lista das palavras-chave dos “ingredientes” nas quatro variantes). A seguir, comentamos brevemente os resultados que mais nos chamaram a atenção:

#### **Português**

- A lista de palavras-chave em POB (60 itens) é cerca de 30% maior que a lista de POP (41 itens). Isso pode ter uma relação direta com o tamanho dos corpora (conforme salientado anteriormente neste artigo, o corpus de POP é o menor de todos em número de palavras). Por outro lado, pode ser um indicativo de que o vocabulário usado na descrição dos ingredientes em POB é mais diversificado, se diferenciando mais claramente daquele usado nas outras seções da receita, se comparado ao vocabulário usado nos ingredientes em POP;
- A lista de POB contém vários participios (que aqui funcionam como adjetivos), indicativos de um preparo prévio do ingrediente (*picada, picado, picados, cortado, picadas, ralada, cortados, cortados, cortadas, amassados, moída, ralado, (em) cubos, cortada, socados*). Esses itens não ocorrem em POP e, assim, poderíamos concluir que: i) o POB é muito mais específico e detalhista na descrição dos ingredientes do que o POP; ii) POP, ao contrário do que se observa em POB, indica o preparo dos ingredientes no “modo de fazer”;
- Outra diferença a ser notada é a maior variedade de itens lexicais usados em POB para indicar unidades/medidas para os ingredientes: [colher de] *chá, sopa, colheres,*

*xícara, colher, xícaras, xíc, café / g, kg, l, ml, litro / copo, lata, pacote, tablete, uni(dades), pitada, ramos, pé, a + gosto. Em POP temos: “sopa, colheres, colher, chá, c, ch, chávena, sobremesa, colh / gr, g, dl, grs, kg / pacote, lata, cálice / dentes / q[quanto] + b[aste]]*

### Inglês

- as duas listas apresentaram um número de itens mais parecido entre si que no caso do português: 82 itens para IOA e 67 para IOB. Esses 20%, aproximadamente, de diferença no tamanho das listas pode ser decorrente do mesmo motivo apontado para o português: o corpus IOA tem mais palavras que o corpus IOB;
- Também no que diz respeito ao tipo de itens lexicais que aparecem nas duas listas de palavras-chave, o inglês apresenta mais semelhanças do que o português. IOA tem 22 participios indicativos de preparo prévio dos ingredientes ao passo que IOB apresenta 16. No caso dos adjetivos qualificativos (como *unsalted, whole*, etc.), IOA tem 13 e IOB 14. A maior diferença observada foi quanto à presença dos advérbios: IOA tem 4 (*finely, thinly, coarsely* e *freshly*) contra apenas 2 (*freshly* e *coarsely*) em IOB, o que pode ser indicativo de maior precisão na variante americana;
- No que diz respeito aos itens lexicais indicativos de pesos, medidas e unidades, a semelhança também ficou evidente: IOA tem 9 itens de medida de precisão (*ounces, lb, oz, pound, ml, pounds, ounce, lbs, fl[uid]*) contra 8 de IOB (*oz, g, pint, ml, lb, fl[uid], pints, kg*). A diferença aparece no uso do sistema métrico internacional, e da unidade de medida “pint” em IOB, ausentes em IOA. As unidades de medida caseira aparecem 11 vezes em IOA e 7 vezes em IOB, sendo possível observar uma diferença tanto na abreviação das duas variantes quanto na presença mais significativa dessas unidades nas primeiras posições da lista IOA, em oposição à presença mais marcante de medidas precisas nas mesmas posições da lista IOB. Os itens *cloves, dash* e *clove* ocorreram em IOA em oposição a *cloves, sprigs* e *bunch*, em IOB.



**C.** o corpus total de cada variante com o total da variante correspondente: **POBtotal** com **POPtotal** e **IOAtotal** com **IOBtotal** (vide ANEXO IV (POB/POP) e ANEXO V (IOA/IOB)). Embora a comparação não seja indicada nos casos em que o corpus de referência não é pelo menos algumas vezes maior que o corpus de estudo, resolvemos testar a metodologia. Os resultados corroboraram a nossa hipótese inicial: a *Keyword* resultante da comparação de POB com POP contém muito mais itens do que aquela resultante da comparação entre IOA e IOB. Isso sugere, pois, que o léxico que distingue POB de POP contém mais itens que aquele que distingue IOA de IOB. A seguir comentamos alguns dos dados que nos chamaram mais a atenção:

- A lista de palavras-chave do português apresenta 80 ocorrências para POP e 63 para POB. Desses itens, 55 (69%) da lista POP têm ocorrência = 0 no corpus de POB, e 22 (35%) da lista de POB tem ocorrência = 0 no corpus POP. Já o inglês apresenta 46 palavras-chave para IOB e 60 para IOA, das quais 17 (37%) só ocorrem em IOB e 27 (45%) só em IOA. Esses números mostram que a diferença entre as variantes do português é mais expressiva do que entre as do inglês, especialmente se levamos em conta que o corpus de POP é o menor de todos (17.661 *types*) e o IOA o maior (26.990);
- Das palavras-chave com ocorrência 0 nas variantes do português destacamos:
  - i) diferenças sintáticas – POP dá preferência pela forma verbal acompanhada da partícula apassivadora “-se” (*deixa-se, junta-se, leva-se, deita-se, serve-se, tempera-se, faz-se, batem-se, coloca-se, mistura-se, retira-se*) enquanto no POB vemos o predomínio do imperativo (*abaixe, abra, doure, desligue*). Neste caso, ainda que esteja presente, em POP percebe-se a preferência por formas verbais pouco ou não usadas em português brasileiro (*deite, tape, coza, aloure, rectifique*);
  - ii) diferenças lexicais – além das apontadas para os verbos acima, encontramos pares POP/POB como os seguintes: *lume / fogo, q.b. / [a] gosto, sumo / suco, confecção / modo (de fazer), tacho – caçarola / panela, chávena / xícara, filetes / filés, aipo / salsão, etc.*;

**iii)** diferenças culturais – para designar a seção da receita em que se explica seu preparo, POP prefere usar *Preparação* (77 ocorrências), ao passo que POB emprega, com mais frequência, *Modo de fazer* (o que pode ser constatado nas 81 ocorrências de *modo*). As 12 ocorrências de *entretanto* em POP nos chamaram a atenção. Procurando verificar as linhas de concordância, constatamos que o termo é usado nos mesmos contextos em que usamos *enquanto isso* em POB:

N	Concordance	File
1	cortado às rodela finas, e deixe refogar suavemente. Entretanto, coza a massa em água temperada com sa	\ppm06_~1.txt
2	cebola picada e o azeite e leva-se ao lume a alourar. Entretanto coze-se os dendéns, assim que estiverem	\ppa09_~1.txt
3	e leve ao forno moderado durante duas horas e meia. Entretanto, deite num tachinho o caldo de aves e as n	\ppa10_~1.txt
4	Maria picada. Forrar a tarteira. Colocar no frigorífico. Entretanto fazer o recheio que é só misturar os ingredi	\pteb06_~1.txt
5	atias finas e demolha-se em água sem desfazer muito. Entretanto frita-se a linguiça juntamente com o alho e	\laco03_~1.txt
6	mal escorrido, e espalha-se por cima a broa migada. Entretanto leva-se o azeite ao lume com os alhos e de	\laco09_~1.txt
7	+ 30 minutos, regando o peixe com o próprio molho. Entretanto leve ao lume num tachinho a manteiga a de	\ppp04_~1.txt
8	inteiros para enfeitar. Deixa-se ferver+ 10 minutos. Entretanto num almofariz, pisam-se os alhos, coentro	\ppp07_~1.txt
9	margarina. Cobrem-se com água e deixam-se ferver. Entretanto pica-se a cebola muito finamente e leva-se	\pien08_~1.txt
10	funda de tamanho médio com o óleo e deixe aquecer. Entretanto, pincele os rolos de frango com a manteiga	\ppa01_~1.txt
11	. Leve a cozer em forno bem quente (200°C a 220°C). Entretanto prepare a calda de açúcar: dissolva o açúc	\pteb08_~1.txt
12	e regue o pato com a calda formada. Deixe cozinhar e entretanto prepare a calda: Leve o açúcar a dourar em	\ppa06_~1.txt

- Nas variantes do inglês, nos chamaram a atenção (IOA/IOB):
  - i) diferenças de grafia: *lasagne / lasagna, centre/ center*;
  - ii) a grande presença das medidas nas duas listas e suas formas variadas de abreviação, em especial as colheres: *TS-TB-T / TBSP* para “tablespoon” e *TEASP* para “teaspoon” em IOB;
  - iii) os pares contrastivos, como *cornstarch / cornflour, shrimp / prawn, broth / stock*;
  - iv) assim como nas variantes do português, em IOA o trecho da receita que descreve seu preparo é predominantemente denominado de *Directions* (42 ocorrências), ao passo que, em IOB, parece haver uma alternância entre o uso de *Method* (61 oc.) e *Instructions* (29);
  - v) a presença de *gas* (47 oc.) + *mark* (26 oc.) e *ml* (47 oc.) – medida do sistema métrico internacional – na lista de palavras-chave de IOB poderia ser um indicio de que, diferentemente dos americanos, os britânicos têm uma preocupação em tornar a receita acessível a um público maior, englobando outras

culturas que não utilizam o sistema métrico anglo-saxão. Interpretamos esse fato como um reflexo de sua condição geopolítica enquanto membro da União Européia e parte do continente Europeu.

Por outro lado, a lista de palavras-chave do IOA ilustra uma tendência cultural crescente nos Estados Unidos. Os itens *carbohydrate*, *fiber*, *cholesterol*, *protein*, *sodium* e *calories* são itens típicos de tabelas de informações nutricionais, mais comuns em receitas dietéticas. No entanto, um dos critérios utilizados na coleta foi não incluir receitas desse tipo, o que mostra que tais informações se tornaram corriqueiras e, acima de tudo, necessárias nesse país, muito provavelmente pelo fato de sua população apresentar sérios problemas de obesidade.

### 2.5 Teste e ampliação do corpus

Segundo Atkins *et al*, “corpus growth and feedback” é a última etapa da construção de um corpus. Para saber se o resultado final do corpus serve aos propósitos imaginados quando de seu planejamento é preciso que ele seja testado e, posteriormente, corrigido e aumentado com base nas observações feitas por usuários capacitados.

Nosso corpus foi utilizado recentemente num trabalho de nossa autoria apresentado no PALC 2003, em Lodz, na Polônia (Tagnin e Teixeira, 2003). O retorno que obtivemos foi de grande importância para que pudéssemos planejar a realização dos ajustes e ampliação necessários, conforme comentamos a seguir.

### 3. Considerações finais

Buscamos mostrar, ao longo deste artigo, as etapas envolvidas no planejamento, coleta e avaliação de corpora técnicos, usando, como exemplo, um corpus comparável multicultural de culinária em inglês americano / britânico e português brasileiro / europeu. O corpus foi construído com o intuito principal de

verificar, por meio do estudo de corpora, as diferenças entre as variantes do português e do inglês (para uma análise mais detalhada dos corpora, vide Tagnin & Teixeira, no prelo). Tais diferenças, a nosso ver, no geral são ignoradas pelas obras de referência e de consulta na área, as quais, inclusive, são tradicionalmente fontes de referência para o tradutor técnico de culinária.

Durante a fase de teste do corpus observamos que, de modo a garantir a representatividade mínima nas quatro variantes, será necessário dobrar o número de receitas coletadas, além de incluir a categoria *Lanches* com suas 20 receitas em cada variante. Ficou clara também a necessidade de tratarmos do aspecto da representatividade no que diz respeito à importância/recorrência específica de cada tipo de receita no contexto gastronômico e cultural de cada variante. Para isso, faz-se necessário adotarmos uma nova postura metodológica e não mais trabalhar com corpora comparáveis no nível das receitas, mas sim no nível cultural. Ou seja, se os pratos com “peixes e frutos do mar” são mais freqüentes em Portugal e na Grã-Bretanha do que no Brasil ou nos Estados Unidos, nosso corpus deverá respeitar e refletir essa discrepância coletando receitas desse tipo mais naquelas variantes do que nessas. Da mesma forma, o alto número de receitas “discursivas” (que não separam “ingredientes” de “modo de fazer”) em POP seria levado em consideração, bem como a representação da diversidade regional de cada variante. Para dar conta dessas diferenças pretendemos coletar 4 sites integrais, um em cada variante, e fazer novas análises.

É importante ressaltar, no entanto, que só pudemos chegar a essas conclusões por meio da construção do corpus, o que evidencia ser a Linguística de Corpus uma metodologia relevante para o tradutor, quer proporcionando-lhe, por meio da construção e análise de um corpus, um processo de aprendizagem que lhe revela características culturais, pragmáticas, sintáticas e lexicais da área em estudo, quer auxiliando-o, durante o processo tradutório, a encontrar, no corpus construído, os equivalentes mais adequados para os termos que procura.

### Referências bibliográficas

- ATKINS, S.; CLEAR, J.; OSTLER, N. (1992) Corpus Design Criteria. *Literary and Linguistic Computing* 7:1. Oxford: Oxford University Press, p. 1-16.
- AUBERT, F. H. (1996) *Introdução à metodologia da pesquisa terminológica bilingüe*. São Paulo, Humanitas/FFLCH-USP.
- \_\_\_\_\_. (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*. Humanitas: São Paulo.
- BAKER, M. (1995) Corpora in Translation Studies: An Overview and Some Suggestions for Future Research. *Target*, 7:2. Amsterdam, p. 223-43.
- BIBER, D. (1993) Representativeness in Corpus Design. *Literary and Linguistic Computing* 8:4. Oxford, Oxford University Press, p. 243-57.
- BIDERMAN, M.T. (1992) O léxico, testemunha de uma cultura. *Actas do XIX Congreso Internacional de Linguística e Fililoxia Romanicas*. A Coruna: Fundación Pedro Barrié de la Maza, Conde de Ferrosa, p. 397-405.
- \_\_\_\_\_. (1996) Léxico e vocabulário fundamental. *ALFA* 40. São Paulo, Unesp, p. 27-46.
- COLINA, S. (1997) Contrastive Rhetoric and Text-Typological Conventions in Translation Teaching. *TARGET vol. 9*. Amsterdam, John Benjamin, p. 335-53.
- CRAZE, R. (1998) *O Guia das Especiarias. Guia do conhecedor*. (sem local de publicação): Livros e Livros. [trad. Jorge Pinheiro Valente].
- FORNARI, C. (2001) *Dicionário-Almanaque de Comers & Bebés*. Rio de Janeiro, Editora Nova Fronteira.
- GOMENSORO, M.L. (1999) *Pequeno Dicionário de Gastronomia*. Rio de Janeiro, Editora Objetiva.
- JAKOBSON, R. (1974) Aspectos lingüísticos da tradução. *Lingüística e Comunicação*. São Paulo, Cultrix.
- JOHANSSON, S. (1980) Corpus-based Studies of British and American English. *Papers from the Scandinavian Symposium on Syntactic Variation*. Stockholm: Almqvist & Wiksell International.
- JUNQUEIRA, L. (1980) *Ervas e especiarias na cozinha*. Rio de Janeiro, Tecnoprint/Ediouro.
- MASSAM, D. e ROBERGE, Y. (1989) Recipe Context Null Objects in English. *Linguistic Inquiry* 20:1, p. 134-9.

TRADTERM, **10**, 2004, p. 313-358

- McENERY, T.; WILSON, A. (1997) *Corpus Linguistics*. Edinburgh: Edinburgh University Press.
- MEIJS, W. (1996) Linguistic Corpora and Lexicography. *Annual Review of Applied Linguistics vol. 16*, p. 99-114.
- NORD, C. (1983) *Text Analysis in Translation*. Amsterdam and Atlanta, GA, Rodopi.
- NORRICK, K. (1983) Recipes as Texts: Technical Language in the Kitchen. In: JUNGEN, R., S. DE KNOP, P.H. NELDE E M.P. QUIX (eds.) *Sprache, Diskurs und Text*. Tübingen: Niemayer, p. 173-82.
- QUIRK, R. (1992) On corpus principles and design. In: SVARTVIK, J. (ed.) *Directions in Corpus Linguistics*. The Hague, Mouton de Gruyten.
- SARDINHA, T.B. (2000) Lingüística de Corpus: Histórico de Problemática. *D.E.L.T.A.*, v. 16, n. 2, p. 323-67.
- TAGNIN, S.E.O. e TEIXEIRA, E.D. (no prelo) British vs. American English, Brazilian vs. European Portuguese - how close or how far apart? - a corpus-driven study. In: LEWANDOWSKA-TOMASZCZYK, B. (ed.) *Proceedings of the 4<sup>th</sup> Conference on Practical Applications in Language Corpora – PALC'03*, Hamburg, Peter Lang.
- TAGNIN, S.E.O. (1989) *Expressões idiomáticas convencionais*. São Paulo, Ática, Série Princípios.
- TEIXEIRA, E.D. (2003) Em busca de um novo modelo tecno-formal para a construção de dicionários técnicos bilíngües – o exemplo da culinária. *Intercâmbio vol. XII*. São Paulo, LAEL – PUC/SP, p. 243-51.
- TEUBERT, W. (2001) Corpus Linguistics and Lexicography. *International Journal of Corpus Linguistics vol. 6* - Special Issue, p. 125-53.
- VAN DEN BROECK, R. (1986) Contrastive Discourse Analysis as a Tool for Interpretation of Shifts in Translated Texts. In: HOUSE, J. e BLUMKULKA, S. (eds). *Interlingual and Intercultural Communication: Discourse and Cognition in Translation and Second Language Acquisition Studies*. Tübingen, Gunter Narr, p. 37-47.
- VÁRADI, T. e KISS, G. (2001) Equivalence and Non-equivalence in Parallel Corpora. *International Journal of Corpus Linguistics vol. 6* - Special Issue, p. 167-77.
- WERLE, L. & J. COX. (2000) *Ingredientes*. Colônia: Könemann. [tradução de Israel Jelin]
- WITTMAN, L.H.; PÊGO, T.R. e SANTOS, D. (1995) Português Brasileiro e Português de Portugal: algumas observações. *XI Encontro Nacional da APL*. Lisboa.