

Linguística de Corpus e Classes de Objeto na Pesquisa Lexicográfica Bilingue

RESUMO

Nosso objeto de estudo, os predicados semânticos da Culinária francesa e brasileira nos ajudaram a pensar num método que abarcasse aspectos linguísticos e estatísticos do domínio da Culinária. A análise contrastiva dos predicados extraídos dos corpora revelou que a complementação do modelo de classes de objeto pela Linguística de Corpus traz maior visibilidade do objeto e, portanto, maior precisão dos resultados.

1. Motivação da pesquisa

Visamos construir um protótipo de glossário bilíngue de verbos de culinária, tendo como língua de partida o francês e de chegada o português, e aplicando conjuntamente os parâmetros de análise de Gaston Gross (1994) – Classes de Objeto – e da Linguística de Corpus (LC). A junção destas vertentes metodológicas surge como uma possibilidade de se obter maior rigor na análise de candidatos a equivalentes. Da LC (BERBER SARDINHA, 2004; TAGNIN & TEIXEIRA, 2004) emprestamos a referência estatística, aliada das buscas por traços definitórios dos elementos analisados, expandindo assim a perspectiva aristotélica de definição (concreto, abstrato, animal, humano, vegetal, etc).

Os dicionários bilíngues, se ainda lacunosos em termos de definição (BÉJOINT & THOIRON (org.), 1996, BONAN-GARRIGUES, 1993), se explicam por conta da comparação de verbetes, tradição metodológica na Lexicografia bilíngue? Esta questão acompanhou todo o desenvolvimento da pesquisa.

2. Revisão bibliográfica

A abordagem de Classes de Objeto tem como objetivo descrever a língua natural por meio das propriedades linguísticas dos seus elementos, cuja estabilidade semântica somente se realiza plenamente na frase. Portanto os predicados compostos por operador e argumento(s) são a unidade elementar nesta perspectiva teórica e não mais unidade lexical. Apoiamo-nos no princípio da regularidade herdado por Harris (1988) e estudados pela LC, para selecionar os candidatos a equivalente, o que contribuiu significativamente para desambiguar unidades lexicais assemelhadas.

Embora o caráter seletivo do predicado não implique em previsibilidade de ocorrência do mesmo, este estudo aponta para a probabilidade da ocorrência existir num determinado domínio.

Por Classes de Objeto entendemos unidade lexical mínima que proporciona estabilidade semântica ao operador ao qual está associada (GROSS, G. 1994).

3. Objetivos específicos

Considerando que o objetivo central é a construção dos equivalentes com base nos parâmetros teóricos inicialmente expostos, nossos objetivos específicos se norteiam pela aplicação destes parâmetros tendo em vista a verificação da eficácia dos mesmos.

Quanto à especificidade dos predicados de um determinado domínio, a título de ilustração, citamos uma rápida pesquisa realizada no corpus de francês revela que os argumentos <gâteau>, <pain>, <brioche> e <pizza>, traduzidos respectivamente por “bolo”, “pão”, “brioche” e “pizza” aparecem associados predominantemente ao operador <cuire>. que nos é apresentado por dicionários bilíngues como equivalente a

cozinhar, como atestamos abaixo, na página de um dicionário didático muito comercializado no Brasil:



Comparando *assar* e *cozinhar*, os verbetes do Dicionário de Usos do Português do Brasil (BORBA, 2002) inserimos as respectivas definições na ficha terminológica abaixo:

Ficha terminológica: verbos *assar* e *cozinhar*.

VERBOS	DEFINIÇÕES
Assar	V [ação processo] [±Compl :nome concreto] 1. cozer a seco sob a ação do calor : <i>As mulheres ainda assavam ao fogo várias frutas, batatas, cocos, etc. [...]</i> 2. abrasar ; queimar : <i>O calor assa. [...]</i> permitir fazer assados: <i>O continental 2001 assa, frita, coze, etc. [processo] [...]</i> 5. tornar-se cozido pela ação do calor : <i>Vai olhar se bolo está assando bem (VD) – DUP, p. 141.</i>
Cozinhar	V [ação processo] [±Compl :nome concreto] 1. preparar ao fogo ou calor ; submeter alimentos ao fogo ou calor ; submeter alimentos ou outra substância à ação do fogo : <i>Aprendi, com Jerônimo, a tirar leite das cabras, a cozinhar o aipim [...]</i> [processo]3. Chegar, sob a ação do fogo ou calor, ao ponto próprio para o consumo; mudar da condição de duro ou cru; amolecer-se ao fogo: <i>Refogue o arroz e deixe cozinhar. – DUP, p. 418</i>

Percebemos que as definições, mesmo num dicionário monolíngue, não apresentam distinções muito evidentes relativas a estes dois verbos. Isto reforça a necessidade do estudo contrastivo dos predicados extraídos de corpus na elaboração de materiais lexicográficos bilíngues.

Ao “garimparmos” os argumentos associados a cada operador e classificando-os em termos de classes de objeto (classes semânticas), visualizamos a seleção que o operador exerce sobre os argumentos. Sabemos que <assar> e <cozinhar> são muito distintos enquanto procedimentos culinários, e tal distinção se evidencia quando associamos <cozinhar> <ovo> e <assar> <ovo> ou ainda <assar> <torta> e <cozinhar> <torta> e o corpus comprova esta “seleção” dos predicados, como vemos abaixo:

Tabela de Classe de Objeto de <assar>

Carnes e Aves	Parte do animal e derivados	Peixe e frutos do mar	Vegetal	Massa	Prato
Frango Leitão/porco Pato Peru Cabrito Coelho	Picanha Pernil Bife Costela Coxa Sobrecoxa Posta Filé	Lagosta Salmão Robalo Tainha Garoupa Haddock Namorado Bacalhau Sardinha	Batata Cebola Cenoura Abacaxi Maça Pêssego Pimentão Abóbora Berinjela	Pizza bolo Pão Rosca Rocambole Torta Quiche Lasanha	Espetinho Rocambole (de carne) Kibe
Frequência e porcentagem 2233 ou 0,062%					

Observamos que o operador <assar> apresenta mais de 2.200 ocorrências no corpus enquanto que <cozinhar> é contabilizado em 8000 ocorrências, aproximadamente. Esta diferença frequencial se sustenta, inclusive, pela diversidade das Classes de Objeto: quanto maior diversidade de argumentos, portanto de Classes de Objeto, maior a probabilidade que o operador ocorra, e o contrário, menor a diversidade de Classes de Objeto, frequência menos expressiva. Desta forma, além das distinções dicionarizadas nos verbetes, podemos verificar o nível de frequência inclusive dos candidatos concorrentes a equivalente. Observemos melhor o operador <cozinhar>.

Tabela de Classe de Objeto de <cozinhar>

Carne Ave Suíno	Parte animal e derivado	Peixe e fruto do mar	Vegetal	Farináceo	Lactíneo	Prato
Frango Porco Vaca Boi	Alcatra Carne moída Lagarto Linguíça Lombo Miúdo Paio Patela Peito Presunto Salsichão Salsicha	Bacalhau Cação Camarão Lula Marisco Polvo Vieiras Vongole	Abobrinha Alho Alho-poró Arroz Aspargo Beterraba Carne de soja Carne vegetal Cebola Chuchu Cogumelo Couve-flor Ervilha Espinafre Feijão Feijão branco Lentilha Mandioca Mandioqui-nha Milho Moranga Morango Tomate	Fusilli Macarrão Nhoque Polenta Ravioli Sêmola Spaghetti	Creme de leite Leite condensado	Caldo Charutinho Creme Cuscuz Geleia Pudim Risotto Sopa Camarão
Frequência e porcentagem		8000 (aprox.)* – 0,233%				
*Foram subtraídas por volta de 200 ocorrências com o substantivo <i>cozinha</i> contabilizadas como verbo pelo concordanciador UNITEX.						

Além das sete Classes de Objeto associadas a <cozinhar> contra seis Classes de Objeto associadas a <assar>, provando que há maior variedade de argumentos entorno de <cozinhar> que de <assar>, confirmamos maior presença do operador <cozinhar> no corpus: 0,233% versus 0,063%.

Quando observamos os operadores <rôtir> e <cuire>, constatamos a mesma predominância de <cuire> e variedade de Classes de Objeto no corpus PT, porém

As buscas no corpus PT revelaram que há 35 ocorrências para <apimentar> apenas, contra 420 do operador do francês, o que nos motivou a buscar a frequência do argumento <pimenta>, para verificar os verbos associados a ele. Vejam no quadro abaixo a frequência obtida:

Figura: Concordanciador Unitex e os predicados com argumento <pimenta>



São mais de 7.400 ocorrências, na maioria das ocorrências, sem operador associado, como este recorte do UNITEX nos mostra. Porém, buscando por <acrescentar> <polvilhar> <adicionar> <colocar> + <pimenta> obtivemos frequências inexpressivas.

Isto equivale a dizer que *apimentar* não é de uso corrente nas receitas brasileiras e que *pimenta*, ingrediente nacionalmente difundido, ocorre em aproximadamente 0,20% do corpus PT, porém, quando associado a verbos, a porcentagem dos predicados decresce para menos de 0,001% - aproximadamente 5 ocorrências com <adicionar> e <acrescentar>.

Outra constatação: nas receitas do Brasil, <pimenta> aparece ao lado de <sal>. Assim, “sal e pimenta a gosto”, por exemplo, cabe como fraseologia candidata à entrada do glossário, em detrimento de *apimentar*. Além do fato de *poivrer* originar-se do termo *poivre*, diverso de *piment*. Temos que, etimologicamente, *poivrer* se refere a acrescentar pimenta do reino e *pimenter* a adicionar *piment*, que não aparece no corpus FR

5. Resultados

Feito o levantamento dos dados, classificamos três tipos de equivalência a relação dos operadores do português e do francês como **equivalência total**, como *mélanger* e *misturar* cujas definições dicionarizadas se assemelham e as ocorrências também, inclusive em termos quantitativos. A **equivalência parcial** ocorre no caso de haver termos equivalentes concorrentes, como *couvrir*, que aceita *cobrir* e *tampar*.

Usando ainda as categorias de Teixeira (2008) sobre equivalência, temos a **não-equivalência** como é o caso de *poivrer*, *équeuter* e *déchiqueter*.

6. Conclusões

A pesquisa comprovou a relevância dos métodos aplicados em objetos da Lexicografia bilíngue, além de abrir possibilidades para mais investigações. As classes de objeto amplia a visibilidade dos eventos linguísticos, oferecendo melhores resultados e mais rigor, enquanto o aparato matemático da Linguística de Corpus evidencia permite sustentar a análise a partir de parâmetros quantitativos.

7. Referência Bibliográfica

BÉJOINT, H. & THOIRON, P. (org), Les Dictionnaires Bilingues, Louvain la Neuve, Duculot. 1996

BERBER SARDINHA, T. Linguística de Corpus. São Paulo: Manole. 2004

BONAN-GARRIGUES, M. Méthode de paramétrage des dictionnaires et grammaires électroniques – Application à des systèmes interactifs en langue naturelle. Tese (Doutorado em Ciência da Linguagem). Département de Recherches Linguistiques – Laboratoire d'Automatique Documentaire et Linguistique. Université de Paris 7. 1993.

BORBA, F. S. Dicionário de Usos do Português do Brasil. São Paulo: Ed. Atica. 2002

DURAN, M. S. Parâmetros para a elaboração de dicionários bilíngues de apoio à escrita em línguas estrangeiras. Tese (doutorado) UNESP, São José do Rio Preto, 2008.

HARRIS, Z. Language and Information. New York: Columbia University Presse. 1988

PAUMIER, Sébastien. Unitex - Manuel d'utilisation. Paris: Univ. Marne-la-Valée, 2004. Disponível em: <http://www-igm.univ-mlv.fr/~unitex/UnitexManual2.1.pdf>. Acesso em: 05/06/2012

TAGNIN, S. E. O.; TEIXEIRA, E. D. Linguística de Corpus e Tradução Técnica – relato da montagem de um corpus multivarietal de culinária. Tradterm, São Paulo, v. 10, n. 1, p. 313-358, 2004.

TEIXEIRA, E. D. A linguística de corpus a serviço do tradutor: proposta de um dicionário de culinária voltado para a produção textual. Tese (Doutorado em Linguística). Faculdade de Filosofia, Letras e Ciências Humanas (FFLCH). São Paulo, 2008. Disponível em: http://www.teses.usp.br/teses/disponiveis/8/8147/tde-16022009-141747/publico/TESE_ELISA_DUARTE_TEIXEIRA.pdf. Acesso em 05/06/2012