# Ficha Terminológica Informatizada: etapas e descrição de um banco de dados terminológico bilíngüe.

Guilherme Fromm[1]

RESUMO: o objetivo deste texto é apresentar um banco de dados, ainda no estágio de desenvolvimento, que será incorporado ao Projeto Comet/USP. Esse banco servirá para a elaboração de fichas terminológicas semi-automatizadas e será alimentado pelos diversos corpora existentes no projeto. Essas fichas prevêem a elaboração de vocabulários técnicos baseados unicamente em corpus.
UNITERMOS: Lingüística de Corpus, Banco de Dados, Terminografia, Terminologia, Tradução.

ABSTRACT: this text aims the linguistic description of a data bank, still being developed, that will be incorporated to the Projeto Comet/USP. This bank is being developed to fulfill semi-automatic terminological cards and will be fed by various corpora available at Comet. These cards preview the construction of technical vocabulary based only on corpora.
KEYWORDS: Corpus Linguistics, Data Bank, Terminography, Terminology, Translation.

O projeto COMET (Corpora Multilíngüe para Ensino e Tradução)[2], da FFLCH/USP, coleta, já há alguns anos, vários corpora em diferentes áreas. Os alunos de mestrado e doutorado da Profa. Dra. Stella E. O. Tagnin, coordenadora do projeto, além de contribuírem para a construção desses corpora, também vêm desenvolvendo trabalhos de pós-graduação para expor diferentes maneiras de trabalhar com esses corpora. Uma das vertentes de estudo é a construção de vocabulários baseados em corpora de áreas de especialidade. Foram tomados como modelos dicionários, monolíngües ou bilíngües, baseados em grandes corpora gerais de língua (como as das editoras inglesas Longman e a Oxford, que trabalharam com corpora próprios ou o British National Corpus). A proposta de alguns doutorandos é a construção de ferramentas e modelos que funcionem como alicerce para a futura organização de obras terminológicas baseadas exclusivamente em corpora.

Qualquer trabalho terminológico pressupõe várias etapas para a construção do produto final, que seria um vocabulário de uma determinada área ou um glossário (usando as concepções de dicionário, vocabulário e glossário, propostas por Barbosa, 2001). Entre essas etapas, uma das mais importantes é a organização dos dados recolhidos através de uma ficha, comumente chamada de ficha terminológica. Cabré (1993) nos explica o que vem a ser essa ficha:

> Las fichas terminológicas son materiales estructurados que deben contener toda la información relevante sobre cada término. Las informaciones que presentan se extraen de las fichas de vaciado o de la documentación de referencia, y se representan siguiendo unos critérios fijados previamente.

[1] FFLCH/USP – UNIBAN.
[2] Uma melhor descrição do projeto pode ser vista em Tagnin, 2005.

> Hay muchos modelos de fichas terminológicas, de acuerdo com los objetivos de cada trabajo y las necesidades de cada organismo. De entrada, podemos distinguir entre fichas monolingües, fichas monolingües com equivalência y fichas bilingües o plurilingües.

A ficha terminológica foi, durante muito tempo, elaborada e preenchida através de um trabalho manual. O advento dos computadores permitiu não só o desenvolvimento da Lingüística de Corpus[3], como também a informatização dessas fichas e a construção de bancos de dados. Propomos, a partir desse momento, a construção de um banco de dados terminológico bilíngüe para o projeto COMET.

O objetivo inicial da construção desse banco é prover o Projeto COMET de uma ferramenta informatizada semi-automática que auxilie no desenvolvimento de obras terminológicas desenvolvidas a partir da grande base de corpora bilíngües já levantadas e disponibilizadas pelo mesmo. O objetivo secundário é fornecer uma base para o desenvolvimento de novas ferramentas ligadas à extração de termos a partir de corpora de áreas de especialidade e a construção de novas ferramentas de visualização do produto final (vocabulários técnicos bilíngües) para diferentes usuários.

Embora existam vários programas disponíveis no mercado internacional (como o Multiterm, Term-PC e outros, muito bem analisados por Gavenski, 2001) e vários bancos de dados terminológicos já desenvolvidos no país, como os pequenos bancos usados pelo CITRAT/CETRAD/USP no ensino de Terminologia para a área de tradução[4] ou os grandes bancos, como o TERMISUL (Maciel, 2001), pensou-se na construção de um banco personalizado para as necessidades do COMET. A vantagem, além do baixo custo de desenvolvimento (a serviço da Empresa Jr., do ICMC/USP São Carlos), é a possibilidade de agregação de novos módulos, associados às pesquisas de mestrado e doutorado sob a responsabilidade da Profa. Stella e outros.

A criação de uma ficha terminológica é essencial para o desenvolvimento de um vocabulário técnico. Vários modelos já foram propostos e, entre eles, podemos citar Aubert (1996), Krieger & Finatto (2004), Gavenski (2001), Bacellar (2002). O modelo que tomamos como ponto de partida para esse banco, no entanto, é baseado em Fromm (2002)[5]. A proposta da dissertação de mestrado do autor era mostrar um modelo para a construção de vocabulário

[3] "A Lingüística de Corpus ocupa-se da coleta e da exploração de corpora, ou conjuntos de dados lingüísticos textuais coletados criteriosamente, com o propósito de servirem para a pesquisa de uma língua ou variedade lingüística. Como tal, dedica-se à exploração da linguagem por meio de evidências empíricas, extraídas por computador". (Sardinha, 2004, p. 3).

[4] Desenvolvidos pelo Prof. Dr. Francis H. Aubert, baseados em um modelo construído no banco de dados Access, da Microsoft.

[5] A ficha terminológica ali apresentada está disponibilizada aqui como anexo.

especializado de informática para tradutores. Usando como base a ficha terminológica monolíngüe não-informatizada ali apresentada, desenvolvemos uma nova proposta para uma ficha monolíngüe com equivalência, que servirá de base para a construção do banco de dados. Em conversas com o técnico da Empresa Jr., decidiu-se pela construção de um banco de dados padrão SQL, com duas tabelas básicas para a inputação de dados. Devido à complexidade de trabalho num banco de dados desse padrão, será criado um ambiente WEB para que os pesquisadores possam preencher as fichas. Em virtude dos custos de elaboração do projeto, somente um administrador terá acesso ao controle do banco numa primeira fase. Ao administrador caberá o cadastro de pesquisadores (para que esses possam alimentar as fichas) e somente ele poderá aprovar as fichas, sendo que essas só serão disponibilizadas para consulta no sistema após aprovação pelo mesmo. Ao administrador caberá, também, a inserção de novas fichas terminológicas, atualização e remoção de fichas existentes no sistema.

A primeira tabela do banco servirá para a inputação de contextos (previamente selecionados) retirados de um corpus de especialidade de uma área escolhida. Serão colocados, para cada termo, tantos contextos quanto os extraídos do corpus e preenchidos os campos relativos a cada um: exemplo, fonte, data de coleta, data de inserção. A partir de cada contexto, o pesquisador deve, também, extrair um conceito do mesmo. Devemos lembrar que ainda na primeira tabela, com a visualização dos contextos em destaque, serão extraídas várias informações morfológicas, sintáticas, semânticas e relativas ao corpus possíveis[6]: entrada equivalente na outra língua, número da acepção[7], posição de freqüência no corpus, formas equivalentes na mesma língua, categoria gramatical, gênero, número, possibilidades de número (para palavras que só existem no singular ou plural), sigla, acrônimo, entrada por extenso, variações morfossintáticas, relações de hiperonímia, relações de hiponímia, relações de co-hiponímia, relações de antonímia, relações de sinonímia, possíveis remissivas. Além disso, o pesquisador poderá cruzar referências com obras já publicadas, verificando se o termo é dicionarizado, se há definições coincidentes, a fonte da definição e a definição dicionarizada em si.

A segunda tabela do banco, disponibilizada numa segunda página de inserção de dados, servirá para a construção da definição do termo. Nela serão visualizados os conceitos

[6] Uma obra terminológica, normalmente, não apresenta aos leitores tantas informações assim. Preferimos, no entanto, elaborar uma ficha com conteúdo o mais abrangente possível, deixando-a mais próxima de uma ficha lexicográfica.

[7] Embora obras terminológicas tendam a apresentar definições monossêmicas, preferimos inserir esse campo. Algumas áreas, que já atualmente apresentam uma grande diversidade de terminologia, como a informática, podem vir a apresentar algumas definições polissêmicas em suas diferentes subáreas.

extraídos pelos pesquisadores na primeira tabela e, a partir dos mesmos, selecionados os traços distintivos. Dali serão tirados o conceito final e a definição do termo[8]. A consulta aos dados do banco poderá ser feita por diferentes ferramentas, que deverão ser desenvolvidas visando à extração de dados específicos ou gerais do mesmo.

Podemos citar, como exemplo de trabalho em curso, a tese de doutorado de Fromm, que proporá um website para o treinamento de alunos de tradução na área de vocabulários técnicos. O usuário final terá acesso aos dados do banco, porém somente para consulta. A inovação proposta será a forma de consultar o banco. A construção do ambiente web está sendo feita em conjunto com a construção do banco de dados. Elisa Duarte Teixeira desenvolve uma pesquisa (ainda em fase inicial), também de doutorado, para a extração de dados diretamente de um corpus, o que providenciará a alimentação automática de exemplos para a ficha terminológica.

**Referências Bibliográficas**


AUBERT, F. H. *Introdução à metodologia da pesquisa terminológica bilíngüe*. São Paulo: Humanitas, 1996.

BACELLAR, F. *Elementos para a elaboração de um dicionário terminológico bilíngüe em Ciências Agrárias*. 2002. 200 f. Tese (Doutorado em Lingüística) – Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2002.

BARBOSA, M. A. Dicionário, vocabulário, glossário: concepções. In: ALVES, I. M. (org.). *A constituição da normalização terminológica no Brasil*. São Paulo: FFLCH/CITRAT, 2001.

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

CABRÉ, M. T. *La terminología*. Teoria, metodologia, aplicaciones. Barcelona: Editorial Antártida/Empúries, 1993. p. 281-282.

FROMM, G. *Proposta para um modelo de glossário de informática para tradutores*. Dissertação (Mestrado em Lingüística). Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2002.

GAVENSKI, M. M. Microisis: uma experiência no gerenciamento de dados terminológicos. In: KRIEGER, M. G.; MACIEL, A. M. B (org.). *Temas de terminologia*. Porto Alegre/São Paulo: Ed. Universidade/UFRGS/Humanitas/USP, 2001.


[8] O conceito final é montado tendo em vista os vários conceitos previamente elaborados pelo terminógrafo. A definição deve obedecer os critérios previamente estabelecidos na construção da obra. Cabré (1993, p. 207-213)

KRIEGER, M. G.; FINATTO, M. J. B. *Introdução à terminologia*: teoria e prática. São Paulo: Contexto, 2004.

MACIEL, A. M. B. Termisul e terminótica. In: *KRIEGER*, M. G.; MACIEL, A. M. B (org.). *Temas de terminologia*. Porto Alegre/São Paulo: Ed.Universidade/UFRGS/Humanitas/USP, 2001.

TAGNIN. S. E. O (2004). Um corpus multilíngüe para ensino e tradução – o Comet: da construção à exploração. *Tradterm 10*. São Paulo: Humanitas, 2004.

| Entrada: Peer-to-peer | Forma Equivalente: P2P, ponto a ponto | Cat. Gram. sm | N° s | Sing./Plural s/pl | Sigla/Acrônimo P2P | Entrada por extenso | Var. Morfossintáticas | Área rede | Acepção n° única | Córpus 1219 |
|---|---|---|---|---|---|---|---|---|---|---|

| | | |
|---|---|---|
| Contexto: Depois de assustar músicos e gravadoras, os programas para troca de arquivos entre internautas (conhecidos como "peer-to-peer") começam a provocar a indústria cinematográfica. A maior parte dos programas "peer-to-peer" que surgiram depois do "Napster", diferentemente da criação de Shawn Fanning, permitem que os usuários compartilhem não apenas vídeos mas também imagens e software, além de músicas em MP3. | Conceito1: programas para troca de arquivos entre internautas | Fonte FSP 11.07.2001 |
| Contexto: Pretendemos desenvolver mais produtos interessantes baseados na tecnologia "peer-to-peer" [ponto a ponto], que permite a comunicação de usuários em locais diferentes" e chamar mais companhias para expandir o negócio. | Conceito2: tecnologia ponto a ponto, que permite a comunicação de usuários em locais diferentes | Fonte FSP 01.08.2001 |
| Contexto: Mas, para fazer tudo isso ao mesmo tempo, só com uma aplicação peer-to-peer (colega a colega ou "entre pares"). Em vez de servidores centralizadores, o peer-to-peer se apóia na conexão direta entre pares, que se revezam nos papéis de cliente e servidor. | Conceito3: conexão direta entre pares (ou colega a colega), que se revezam nos papéis de cliente e servidor. | Fonte OESP 05.03.2001 |
| Contexto: Peer-to-peer - Modelo de conexão no qual cada um dos equipamentos conectados tem os mesmos recursos e cada parte pode dar início a uma sessão. Na internet, refere-se a uma rede transitória que garante a um grupo de usuários com o mesmo programa acessar arquivos instalados no disco rígido de outros. É o princípio utilizado pelo Napster e programas e serviços similares. ... Discussões sobre direito autoral à parte, a verdade é que o P2P nunca mais será o mesmo | Conceito4: rede transitória que garante a um grupo de usuários com o mesmo programa acessar arquivos instalados no disco rígido de outros. | Fonte INFO 02.2001 |

| Con-ceito | Traços Distintivos | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | A | B | C | D | E | F | G | H | I | J |
| 1 | programa | | troca | arquivos | internautas | | | | | |
| 2 | tecnologia | ponto a ponto | comunicação | | usuários | locais diferentes | | | | |
| 3 | | | Conexão direta | | Pares, colega a colega | | Revezamento | Cliente | Servidor | |
| 4 | Rede transitória | | | Acesso de arquivos | Usuários | | | | | Disco rígido |

| | | |
|---|---|---|
| Conceito final: programas que criam uma rede transitória ponto a ponto para a troca de arquivos entre usuários de locais diferentes, revezando as máquinas desses usuários como clientes e servidores e possibilitando o acesso a arquivos nos respectivos diretórios. | Termo Dicionarizado? ( )sim ( x )não<br>Definições coincidentes? ( )sim ( )não ( ) parcial<br>Fonte (s): | Definição Dicionarizada: |