

ULTRAPASSANDO TERMOS E DICIONÁRIOS: A PESQUISA TERMINOLÓGICA COMO ELEMENTO DE DIÁLOGOS TRANSDISCIPLINARES

Maria José Bocorny Finatto

Universidade Federal do Rio Grande do Sul (RS, Brasil)
Projeto Termisul e Projeto TEXTQUIM
mfinatto@terra.com.br

Viviane Possamai

Universidade Federal do Rio Grande do Sul (RS, Brasil)
doutoranda em Terminologia do Programa de Pós-Graduação em Letras.
vivianepossamai@yahoo.com.br

1. Introdução

Este trabalho visa discutir o estatuto atual dos objetos e das ferramentas teóricas em Terminologia tomando como ponto de partida resultados de uma pesquisa concreta particular. A experiência de pesquisa em foco intitula-se “Projeto TextQuim” e está em desenvolvimento desde 2002. De 2003 até 2006, temos contado com apoio de infra-estrutura do grupo TERMISUL, da área de Educação Química da nossa universidade e com o apoio financeiro do CNPq e da FAPERGS, instituições brasileiras de fomento à pesquisa acadêmica. Essas instituições nos concederam recursos para equipamentos, serviços e bolsas para estudantes de graduação que atuam na modalidade Iniciação Científica.

Em primeiro lugar, apresentamos a origem e o desenvolvimento da pesquisa TextQuim, sua natureza, tópicos de estudo e interesses. Depois, apresentamos um dos seus desdobramentos atuais, uma investigação de doutoramento dedicada ao reconhecimento de informações relevantes em textos do tipo *paper* que serão vertidos do português para o inglês. Esse doutoramento segue linha desenvolvida em estudo de mestrado (Possamai, 2004). Por fim, retomamos alguns questionamentos colocados e destacamos algumas peculiaridades da pesquisa terminológica que ultrapassa o foco tradicionalmente restrito à unidade de significação especializada e não envolve produzir glossários ou dicionários especializados. Afinal, seu maior interesse é empreender um reconhecimento amplo de modos de representação do conhecimento via linguagem. Para encerrar, trazemos algumas reflexões sobre os encaminhamentos teórico-metodológicos implicados nesse tipo de pesquisa.

2. O projeto TextQuim

São os seguintes os objetivos colocados para a nossa pesquisa que tem se concentrado no exame da linguagem de Química em língua portuguesa:

Objetivos gerais

- Aprofundar estudo quantitativo e qualitativo do perfil do texto científico de Química escrito em português, continuando parceria com a área de Educação Química da UFRGS (vide www.iq.ufrgs.br/aeq), tomando como referencial teorias de Terminologia e Teorias de Texto e de Discurso para observação de elementos terminológicos e de recursos lingüísticos

- constitutivos de coesão (estruturação sintática), identificando, a partir deles, marcas de enunciação (estruturação semântica);
- Gerar caracterizações lingüístico-textuais detalhadas para apoiar a necessária crítica do material de bibliográfico geral e instrucional de Química, de modo a colaborar para o aperfeiçoamento das suas condições de redação, o que naturalmente origina subsídios para a crítica das condições de representação e de construção de conhecimentos na matéria, vinculando-se este projeto de pesquisa à temática da pesquisa sobre linguagens, leitura e sobre o ensino de ciências.

Objetivos específicos

- Descrever e analisar, do ponto de vista lingüístico-terminológico e enunciativo, formulações causais, modalizações e terminologias presentes em textos científicos de Química, especialmente manuais acadêmicos utilizados durante a formação de Licenciandos em Química, visando contribuir tanto para os Estudos de Texto (Teorias de Lingüísticas dedicadas ao Texto) quanto para os estudos de Educação Química, pesquisas sobre problemas do ensino e aprendizagem dessa ciência.

Para alcançar esses objetivos (gerais e específico), já dispúnhamos de um bom acervo digitalizado de textos de Química (vide www.ufrgs.br/textquim), de diferentes graus de complexidade, distintos níveis de densidade terminológica e dirigidos a públicos especializados variados. O acervo foi organizado entre 2003 e 2005 e atualizado em 2006.

Assim, além de observar modos de expressão de causalidade, de modalização e a presença de determinadas terminologias da Química em um determinado grupo textual, buscamos contrastar seus padrões em textos de diferentes tipos: artigos de periódicos e revistas de popularização de ciência. Os sujeitos-autores dos textos mais explorados, manuais didáticos do ensino superior, são usualmente renomados pesquisadores e/ou professores de Química; seus interlocutores são químicos iniciantes ou estudantes de engenharia e áreas afins, fundamentalmente estudantes de universitários que fazem estudos básicos de Química.

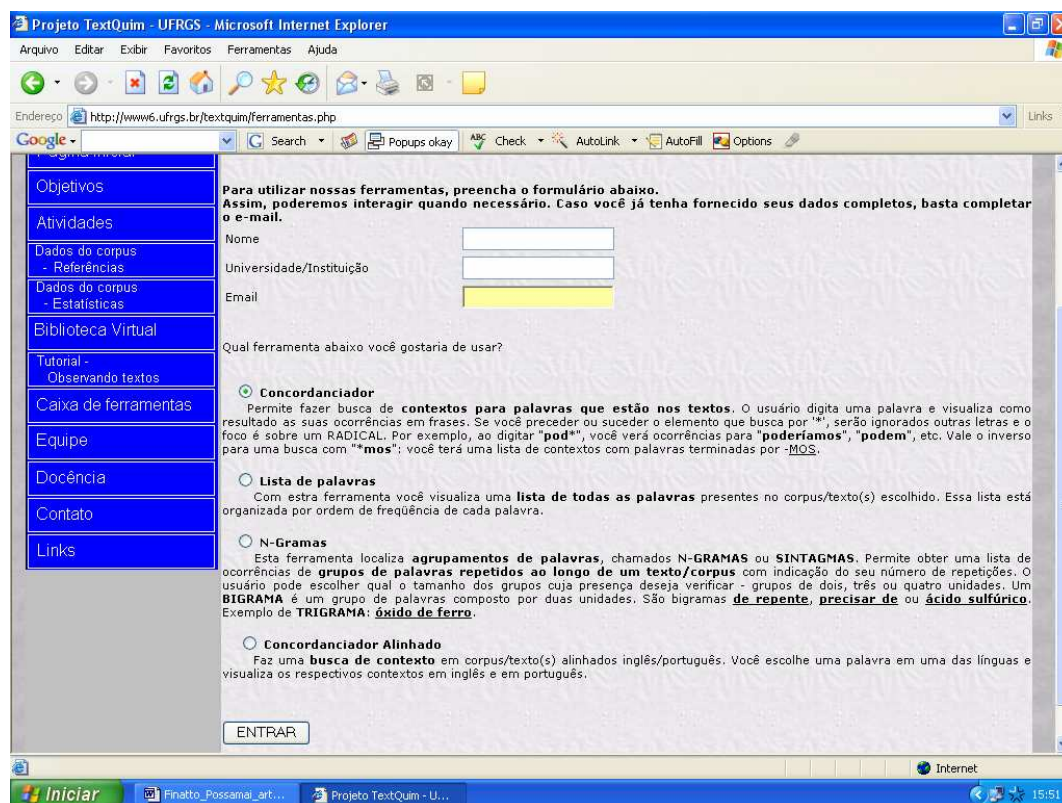
O *corpus* TextQuim tem sido tratado com apoio informatizado, com especial destaque para manuais acadêmico-didáticos dedicados aos macrotemas Química Geral e Físico-Química. Nesse segmento, dispomos de um acervo de textos alinhados, que são originais em inglês e suas respectivas traduções para o português, organizados parágrafo a parágrafo, salientando-se que a grande maioria da bibliografia acadêmica utilizada no Brasil é fruto de tradução, especialmente do inglês. Para maiores detalhes sobre resultados já obtidos, incluindo dados estatísticos, conferir o último relatório da pesquisa disponível no *site* TextQuim antes citado.

A tabela a seguir detalha a composição total do *corpus*, que inclui uma gama de textos de Química em português, inglês e espanhol, cobrindo diferentes gêneros textuais: a) *papers* em inglês, português e espanhol; b) material didático do ensino superior em português e em inglês (Manuais); c) material didático do ensino médio ou colegial norte-americano em inglês (*Chemcom*); e d) artigos de revistas brasileiras de popularização de temas de ciências (Revista Superinteressante).

TIPO DE TEXTO E LÍNGUA	NÚMERO DE PALAVRAS DO CORPUS (tokens)	TIPOS DE PALAVRAS DIFERENTES (types)	RELAÇÃO token/type
Manuais Química Geral / Português	343.279	10.631	3,10
Manuais Físico-Química / Português	89.067	4.687	5,26
Manuais Química Geral / Inglês	168.598	5.887	3,49
Manual Chemcom / Inglês	6.642	1.663	25,04
Artigos revista Química Nova / Português	428.102	25.279	5,90
Artigos revista Química Nova na Escola / Português	59.982	8.705	14,51
Artigos revista Chemistry Education Research and Practice Inglês	15.725	2.274	14,46
Artigos revista Educación Química /Espanhol	65.203	7.948	12,19
Artigos revista Superinteressante/Português	59.551	9.683	16,26
TOTAL corpus inglês + espanhol + português	1.236.149	76.757	não se aplica

A variedade do nosso acervo e a diferença de sua riqueza lexical pelos segmentos que o integram visam representar a diversidade dos textos de Química utilizados e produzidos no Brasil, com ênfase para o texto presente no meio universitário.

Recentemente, passamos a oferecer as seguintes possibilidades de acesso *on-line* a esse *corpus*: a) consulta sobre ocorrências de palavras ou de expressões em diferentes idiomas além do português (inglês e espanhol); b) lista geral de palavras de compõem um texto determinado; c) identificação de expressões co-ocorrentes; e d) acesso a *corpus* alinhado português-inglês. Uma amostra desses recursos pode ser visualizada na figura a seguir.



De modo sintético, podemos dizer que é uma pesquisa dedicada ao reconhecimento da linguagem de Química em suas diferentes modalidades, sem o intuito de gerar um glossário ou de escrever uma terminologia *stricto sensu*, ainda que nos interessem muitas expressões que são conceitualmente relevantes, geralmente equivalentes às terminologias características dessa ciência. O tópico principal é o reconhecimento de um *modo de dizer em ciências* que será verificado em diferentes instâncias discursivas e comunicativas, envolvendo textos de diferentes níveis de especialização e dirigidos a públicos distintos, reunidos em *corpora*. Nessa direção, aliamos-nos à pesquisa em Terminologia, Linguística de Corpus (Berber Sardinha, 2004) e aos Estudos do Texto e do Discurso, com especial destaque para a concepção enunciativa da linguagem (Flores; Teixeira; 2005). Afinal, interessa reconhecer tanto elementos da estrutura dos textos, dos enunciados, quanto de uma enunciação científica.

3. Desdobramentos: da retórica científica à relevância de informações

No “cenário” TextQuim, lidamos com o perfil da linguagem e do texto de uma determinada área de saber. Nele há um espaço especial para abrigar novas pesquisas com foco em aspectos textuais outros que incluam, mas que também ultrapassem, aqueles usualmente recortados em Terminologia aplicada e teórica, quais sejam, termos, definições e fraseologias.

Interesses em aplicações mais abrangentes para os resultados de estudos com textos de áreas técnicas e científicas nos levam a incluir um olhar sobre o texto que, poderíamos dizer, coloca *a priori* todos os itens lexicais de um texto em um plano de igualdade e, depois, “chama” para uma observação em detalhe aqueles que mais possam interessar de acordo com objetivos, finalidades e aplicações de cada item de pesquisa que se enseja.

Próxima a esse “espírito investigativo”, já estava uma pesquisa antes desenvolvida, embora não associada diretamente ao TextQuim (POSSAMAI, 2004). Nela foram levantadas, também a partir de análise de *corpus*, estruturas marcadoras da retórica de artigos científicos de Ciência da Computação escritos em português e em inglês. Essas estruturas foram agrupadas em três categorias, que também foram subdivididas ainda que, por vezes, possam se sobrepor. As categorias identificadas com determinadas expressões foram as seguintes:

Categoria 1

Marcadores relacionados ao propósito do *paper*

Subcategorias

- método (através de um processo, com o auxílio de)
- ancoragem (a partir dos resultados, serve como ponto de partida)
- quantificação (na maioria dos casos, um grande número de)
- tempo (até o presente momento, ao longo do processo)
- explicação (que por sua vez, como por exemplo)
- restrição (no que concerne, de acordo com)
- palavra-núcleo objetivo (com o objetivo de, o objetivo deste artigo é)
- palavra-núcleo forma (da mesma forma que, de tal forma que)

Categoria 2

Marcadores metarreferenciais

Subcategorias

- organização (este artigo está organizado da seguinte forma, a seção # apresenta)
- palavra-núcleo seção (a seção # apresenta, na seção # apresentamos)
- palavra-núcleo figura (como apresentado na figura, x apresentado na figura #)
- palavra-núcleo tabela (a tabela # mostra, como mostra a tabela #)
- palavra-núcleo artigo (este artigo apresenta, neste artigo discutimos)
- palavra-núcleo trabalho (o objetivo deste trabalho é, esse trabalho tem por objetivo)

Categoria 3

Marcadores subjetivos:

- palavra-núcleo importante (*é importante ressaltar que*)
- palavra-núcleo necessário (*é necessário que*)
- palavra-núcleo esperar/acreditar (*espera-se que, acreditamos que*)
- palavra-núcleo observar (podemos observar que, nota-se que)
- palavra-núcleo poder (pode ser definido como, pode-se dizer que)

Observando esses exemplos, podemos concluir, em um primeiro momento, que tais *marcadores* não estão diretamente relacionados com a Ciência da Computação, mas, sim, com a retórica e estruturação do gênero textual artigo científico.

Se olharmos, no entanto, esse tipo de estruturas em artigos de outra área, a Química, por exemplo, teremos uma outra dimensão do caráter “especializado” desses elementos da linguagem. Uma observação de textos produzidos por médicos, por exemplo, levou-nos a um reconhecimento de estruturas relacionadas ao gênero *paper* que resulta em algumas dessas unidades já identificadas e outras, tais como *os dados foram coletados no período de, de acordo com várias pesquisas, nos últimos # meses*. Essas construções, entre outras, a partir de uma observação extensiva, parecem típicas de *papers* brasileiros de Medicina.

Dessa observação, podemos já notar que a separação entre o que é especializado, ou seja, vinculado a uma área e, por extensão, *estritamente terminológico*, e o que é da retórica “discursiva” do texto são aspectos que se misturam, assim como outros itens textuais, tais como elementos coesivos, adjetivos, verbos, palavras do léxico geral, agrupamentos diversos. Assim, todos esses itens formam um todo que é o *paper*, o que corrobora a idéia de que o objeto em foco é o texto e não seus “termos técnicos” isolados.

Dessa experiência de reconhecimento da retórica científica, nasceu uma nova pesquisa, agora de doutoramento, da mesma autora, abrigada pelo Projeto TextQuim, que explora a idéia de localização de *informação relevante* em *papers*. Desta vez, como na anterior, não se pretende explorar apenas termos ou fraseologias e colocações especializados, mas, sim, propor um recorte maior para *tratamento* da informação textual. *Tratamento* implica aqui processamento informatizado, que será gerado via descrição de um *corpus* que alimentará o planejamento de funções de um aplicativo de computador.

A noção *informação relevante* estará vinculada ao objetivo do leitor que utiliza o texto, ressaltando-se um caráter subjetivo da noção. É uma relevância que está mais para o sentido do vetor observador-texto do que o contrário, que é o que parece ser mais comum na pesquisa terminológica, com a incessante e usual busca por definir e recortar do texto um termo X, sintagmas terminológicos Y, o que sempre se mostra tarefa difícil, principalmente em se tratando de unidades fraseológicas especializadas.

Nessa nova iniciativa, sem desprezar o mérito da pesquisa terminológica *stricto sensu*, pressupomos que o observador ou leitor não seja um indivíduo, mas um coletivo, ou uma comunidade, especificamente a comunidade de indivíduos com interesse na tradução dos textos em português para a língua inglesa. Esses dados são importantes, pois, como dissemos, a definição do que será marcado como relevante está pontuada pelo objetivo ou “olhar” desses indivíduos, que têm interesses, dificuldades e expectativas diferentes de, por exemplo, um terminólogo, mas que também trabalham num âmbito de interface com a Terminologia.

Em uma observações piloto de artigos científicos nas áreas de Computação e Medicina, já constatamos que o estabelecimento de um conceito de informações relevantes e a definição de que informações seriam essas não pode ser levado da mesma forma como geralmente conduzimos pesquisas terminológicas, extraindo unidades dos textos. Isso se deve justamente ao paradoxo de como selecionar algumas estruturas como mais relevantes se estamos trabalhando com a noção de informação textual como um todo. Essa perspectiva abre um outro desafio de pesquisa.

Considerando que o objetivo final desse trabalho será o de ajudar o tradutor a encontrar material de referência ou equivalentes na língua para a qual traduz, percebemos que trabalhar com blocos maiores de texto, como parágrafos, poderia apresentar resultados mais positivos. Para a seleção de frases ou parágrafos relevantes, a idéia é identificar palavras-chave, usando o conceito de um *todo de informação aglutinado em torno de algumas palavras* que “magnetizam” informações relevantes e uma determinada função dentro do texto.

Assim, palavras do léxico geral como *abordagem* (e outras, tais como, *classificação, causa, estudo, etc.*), que seriam predefinidas com métodos estatísticos da Linguística de Corpus, poderiam ser usadas para identificar automaticamente sentenças como:

Este artigo apresenta uma **abordagem** para reconhecimento e **recuperação de imagens** estáticas do tipo bitmap baseado em atributos de cor utilizando redes neurais artificiais...

Cruzando-se a palavra *approach* e o termo *image retrieval*, por exemplo, em uma busca no buscador *Google*, teríamos como primeiro resultado uma página na *web* de um projeto sobre extração de imagens, que poderia ser fonte de informação relevante para o tradutor. (<http://wang.ist.psu.edu/image/>). Palavras associadas à informação relevante, segundo palnejamos, poderiam estar todas armazenadas em um dicionário eletrônico de uma ferramenta informatizada, com seus possíveis equivalentes em freqüência e função nas duas línguas.

Obviamente esse método seria um exemplo muito singelo, embora possa ser tido como visionário, mas buscamos o desenvolvimento de um sistema com outras regras e outras associações que permitirão encontrar um resultado mais exato. Esse sistema informatizado deve permitir oferecer ao profissional de tradução uma estrutura frasal muito semelhante à estrutura real da língua de chegada por tipos ou gêneros de textos especializados.

O ponto em que queremos chegar com esse tipo de pesquisa e reflexão é o que, esperamos, permitirá que nos posicionemos sobre a seguinte pergunta: Até que ponto uma avaliação de estruturas como marcadores textuais de gêneros não pode ser considerada uma pesquisa em Terminologia? E, ao avaliar o rendimento de pesquisas atuais e futuras, como a desenvolvida no Projeto TextQuim, pretendemos poder mensurar qual a perspectiva da pesquisa terminológica rumo ao texto e à comunicação, para além dos termos e dos dicionários terminológicos.

Dentro dos exemplos fornecidos acima, perguntamos, ainda: como a Terminologia lidará com essas estruturas lingüísticas e informacionais ao reconhecer que não estão distantes e nem deixam de ser motivadas pela área de conhecimento? Como poderíamos imaginar, então, que um marcador de relevância do tipo X não pode ser tratado ou contemplado em uma pesquisa terminológica estrita?

Na pesquisa sobre identificação de informações relevantes, estamos em busca de elementos que permitam o desenho ferramentas terminológicas para auxílio ao tradutor, ferramentas ou *softwares* que tragam resultados mais completos e contextualizados, economizando tempo de buscas e sendo mais eficazes.

Nessa busca, deparamo-nos com mais uma pergunta entre outras tantas: são as investigações que se ocupam de “mapear” informação relevante para um determinado segmento de usuários-leitores ainda *em e de Terminologia?*

4. Peculiaridades da pesquisa em questão e algumas reflexões

Embora tenhamos colocado, há pouco, uma série de perguntas sem respostas, acreditamos que são pertinentes para uma reflexão em meio a um evento tão importante quanto é o X Simpósio da Riterm. Esse tipo de pesquisa implica uma perspectiva com a qual não estamos, em geral, muito acostumados, visto de ultrapassa o foco mais restrito e usual sobre as terminologias. Esse “novo” foco já foi preconizado por Hoffmann (1982 e outros anos), além de tantos outros autores, de modo que o estatuto de novidade pode ser também questionado. Nessa perspectiva diferenciada, escopo de uma *Linguística do Texto Especializado*, o ponto de maior de atenção é o texto e a linguagem em um uso diferenciado que a singulariza frente a outros usos. Isso, entretanto, instaura um objeto de investigação multifacetado, passível de ser abordado por inúmeros diferentes ângulos, sem qualquer risco de esgotamento.

É nessa linha de raciocínio que entendemos que a pesquisa em Terminologia tem, sim, pleno potencial para incluir aspectos textuais e se fortalecer. Assim como para o desenvolvimento da pesquisa sobre nossos manuais didáticos de Química foi relevante observar, em um dado momento, o comportamento de modalizações e terminologias do ponto de vista enunciativo, entendemos que, para o desenvolvimento da pesquisa terminológica cujos interesses são motivados por fins específicos, suporte à tradução, por exemplo, a observação de trechos maiores de textos e também a observação de *todos* macroestruturais de textos pode proporcionar resultados bastante positivos.

Nessa “inclusão” há pouco mencionada é que queremos nos empenhar para que a nossa área de estudos, que é uma área em franca expansão no panorama acadêmico brasileiro, deixe de ser tida como algo absolutamente recortado e categórico em meio às diferentes tendências mais consolidadas e transdisciplinares dos estudos linguísticos da atualidade. A Terminologia estuda a comunicação especializada e lhe é inerente aproveitar as mais variadas perspectivas e interfaces dos estudos linguísticos, pois nos interessam tanto o todo das linguagens técnico-científicas quanto suas partes, mecanismos e propriedades.

À medida que a pesquisa em Terminologia passa a ocupar-se também dos modos de estruturação do conhecimento científico ou tecnológico, principalmente em iniciativas com parceria ativa e intensa de especialistas que não são linguistas, instauram-se, geralmente por parte desses colegas, várias críticas e questionamentos sobre orientações teóricas e padrões da nossa pesquisa e suas aplicações. Desse diálogo e da parceria com profissionais, tais como médicos, informatas ou engenheiros, entre tantos outros, surge a oportunidade de um auto-questionamento para o terminólogo-lingüista.

Em um cenário de trocas trans- e interdisciplinares, que importância ainda teria uma distinção absoluta entre *termos* e *palavras*? Para contextualizar essa questão derradeira, finalizamos esta apresentação dizendo que cada área de estudos tem seus objetos e é natural que seja assim. Todavia, é vital não isolar esses objetos de um todo sistemático no qual geralmente estão inseridos, de modo que não possamos nos comunicar, por exemplo, com colegas que, também em exemplo, não

tenham textos como objetos do seu estudo ou atenção, mas sim *corpora* e que acreditem que nada na língua é especial ou diferenciado *a priori*, nem mesmo as terminologias.

Bibliografia citada

BERBER SARDINHA, Tony. *Linguística de Corpus*. Barueri-SP: Manole, 2004.

FLORES, Valdir do N.; TEIXEIRA, Marlene. *Introdução à Linguística da Enunciação*. São Paulo: Contexto, 2005.

HOFFMANN, Lothar. Probleme und Methoden der Fachsprachenforschung. *Wiss.Zs. der Karl-Marx-Univ. Leipzig. Gesellschafts- und sprachwiss.* Reihe 31, p.25-34, 1982.

POSSAMAI, Viviane. *Marcadores textuais do artigo científico em comparação português-inglês – um estudo sob a perspectiva da tradução*. Porto Alegre: Universidade Federal do Rio Grande do Sul/Programa de Pós-Graduação em Letras da UFRGS, Dissertação de Mestrado, 2004.