

Terminologia em agropecuária: elaboração de glossários no par linguístico português-espanhol, voltados para o ensino da tradução técnica

Francisca Pereira da Silva - IFRR

Introdução

A pesquisa intitulada “Terminologia em Agropecuária: elaboração de glossários no par linguístico português-espanhol, voltados para o ensino da tradução técnica” teve início com algumas indagações acerca de determinados termos específicos em textos técnicos utilizados no processo de tradução pelos professores da área agrícola do Instituto Federal de Roraima. Algumas dúvidas encontradas no processo de tradução, os mesmos recorriam para buscar o significado nos dicionários ou com a professora de espanhol, mas nem sempre sanavam as dúvidas. Ao se deparar com tais dificuldades, os professores muitas vezes paravam de traduzir e buscavam textos em língua portuguesa a respeito do que queriam trabalhar com os alunos.

Uma boa parte dos dicionários bilíngües especializados disponíveis atualmente em português brasileiro, salvo algumas exceções, consiste de meras listas de termos, muito deles monovocabulares, por vez recopilada a partir de outras listas semelhantes, baseadas em informações desatualizadas e de confiabilidade duvidosa. (Teixeira, 2003 e 2004).

A partir desta inquietação, dentro de um contexto de escola agrícola, onde a prática de utilização de diversos textos é constante, suscita então, a vontade de propor a construção de glossários específicos para estes termos técnicos na área de agropecuária, no qual de conta não só do significado, mas das relações que são estabelecidas no sentido micro do texto.

Para isto, é necessário buscar subsídios teóricos na Lingüística de Corpus, Terminologia e as relações estabelecidas dentro do texto, de modo que possa entrelaçar as teorias para melhor compreensão dos textos no processo de tradução.

O contexto da pesquisa

Percebem-se em textos técnicos específicos de uma área a presença de alguns termos que passam a ser entendidos a partir do momento em que são

analisadas as regularidades, realizações do léxico e sua associação com diversas redes gramaticais mais freqüentes. Assim, a proposta da pesquisa é com o foco em um dicionário, não com conteúdo específico de uma área, mas a forma como os textos veiculam este conteúdo se organiza linguisticamente.

Para isto, ainda que se reconheça o significado de tais palavras, a observação de análise desta pesquisa está centrada em como, quando e quanto elas ocorrem em textos da área de agropecuária, com que palavras estabelecem relações, os tipos de sintagmas e suas características semânticas dentro dos textos, cuja problemática encontra-se baseada na Terminologia.

Considerações teóricas

Para obter estes dados em textos técnicos específicos em agropecuária, buscam-se também subsídios na Linguística de Corpus que “trabalha dentro de um quadro conceitual formado por abordagem empirista e uma visão da linguagem como sistema probabilístico.” (Berber Sardinha: 2004:30). Este tipo de visão é organizado a partir da área específica em agropecuária, com textos (português e espanhol), distribuídos em forma de corpus eletrônico. Há ainda, uma estreita relação da Linguística de corpus e a Terminologia, pois ambas trabalham com o uso de corpora eletrônicos que são utilizados para o levantamento de dados.

Segundo Teixeira (2008) defende a idéia de que os textos são compostos de palavras, em si e individualmente (ou em estado de dicionário) não detém mais sentido que aquele que acumularam ao longo da vida discursiva. Ou seja, seu sentido se reconstrói no texto, de acordo com as relações de significado das palavras que estão ali (e com as que não estão). A linguagem técnica representa nesse sentido, uma forma típica de organização de certas palavras dentro do texto. E nesta organização que os termos técnicos em agropecuária serão trabalhados.

Logo, surge então a pergunta norteadora: No processo de tradução, qual o uso mais freqüente de determinados termos específicos da área? E as subperguntas: Em quais redes gramaticais, lexicais e semânticas estão associadas? Em que material de consulta o tradutor encontra informações para que possa traduzir um texto específico em agropecuária?

Assim, o corpus é um conjunto de textos reunidos no computador, com critérios estabelecidos de modo que possam ser representativo e ficam disponíveis para

serem pesquisados com detalhe ou estatisticamente. Para isto, necessita-se de um programa específico para análise chamado WordSmith. O objetivo desta pesquisa é reunir estes textos específicos da área de agropecuária, tanto na língua portuguesa quanto na língua espanhola para que seja feita uma análise à luz da lingüística de corpus e terminologia acerca destes textos.

De acordo com a definição de Stella Tagnin (2004), um corpus é “uma coletânea de textos em formato eletrônico, compilada segundo critérios específicos, considerada representativa de uma língua (ou da parte que se pretende estudar), destinada à pesquisa”. Algumas ferramentas de análise de corpus concentram-se especialmente na análise sintática, como o *Unitex*, ou na análise lexical, como o *WordSmith Tools*, ao qual voltaremos mais adiante.

Alguns pontos específicos analisados com essas ferramentas são a tradução de fraseologias, colocações e terminologia técnica, por exemplo, que podem ser amplamente beneficiadas pela análise de corpora em língua de origem e língua de chegada, contribuindo, assim, para a produção de um texto mais fluente e natural e com terminologia mais precisa.

O programa *WordSmith Tools*

Em seu estudo de 2002, Tagnin define o *WordSmith Tools* (ou WS Tools) da seguinte forma:

...WordSmith Tools, ferramenta que fornece, a partir de textos pré-selecionados, concordâncias para a palavra de busca, clusters (agrupamentos freqüentes), listas das palavras mais freqüentes num texto, bem como palavras-chave de um texto.

A ferramenta, desenvolvida por Mike Scott, da universidade de Liverpool, é dedicada a análises lexicais, como colocações, agrupamentos e dados estatísticos relacionados a esses e outros itens. Dadas essas características, a ferramenta apresenta-se como sendo de grande utilidade para desenvolvimento de glossários de áreas de especialidade, como aqueles mencionados no estudo de Tagnin.

No âmbito acadêmico, ela vem sendo utilizada em pesquisas das mais variadas áreas, como análise do discurso, sociolingüística, estudos de tradução, incluindo estudos sobre auto-tradução e ensino de tradução, por pesquisadores como Stella Tagnin, FrancisHenrik Aubert e Tony Berber Sardinha.

Os princípios abstratos da ferramenta são:

Ocorrência: somente são considerados os itens que constam no corpus.

Recorrência: deve haver pelo menos duas ocorrências do item. Esse aspecto realça a importância do tamanho do corpus: quanto maior ele for, maior a probabilidade de ocorrência de termos raros.

Concorrência: os itens devem ocorrer junto com outros. O pressuposto teórico aqui é de que o significado de um termo é determinado em relação aos termos com os quais se relaciona.

A próxima seção contém uma apresentação resumida dos principais recursos do WS Tools, com base no livro *Linguística de Corpus* (Sardinha, 2004). É possível fazer download gratuito da versão de demonstração da ferramenta no endereço eletrônico <http://www.oup.com/elt/global/catalogue/multimedia/demo/>. O site do projeto COMET - Corpus Multilíngüe para Ensino e Tradução, da USP, apresenta um manual resumido em português (<http://www.flch.usp.br/dlm/comet/>).

Recursos do *WordSmith Tools*

A ferramenta apresenta três recursos: *WordList*, *KeyWords* e *Concord*.

A *WordList* é utilizada para gerar listas de palavras a partir do corpus de estudo. Uma vez coletado o corpus, de acordo com critérios rigorosos (detalhadamente apresentados em Sardinha, 2004), seus arquivos serão selecionados para a criação da lista.

Na *WordList* são apresentadas todas as palavras do corpus, individualmente, com suas frequências. São geradas duas listas: uma em ordem alfabética e outra em ordem de frequência.

Para evitar a apresentação de várias entradas com o mesmo lema, é possível fazer a lematização: clique no item da lista que será o lema (por exemplo, *casa*), pressione F5 (mark/unmark) e clique nas outras formas que serão relacionadas ao tema (por exemplo, *casamento*, *casar-se*, *casou casado*). Em seguida, clique em *join*. As palavras ficam com a cor cinza, passam a figurar ao lado do lema na coluna *Lemmas* e suas frequências são alteradas de acordo com as novas relações estabelecidas.

A lista de estatísticas apresenta os *tokens* (correspondente ao número total de itens ou palavras, incluindo as repetições de um mesmo item ou palavra), os *types* (correspondente a cada item ou palavra, sem considerar as repetições), a *type-token ratio*, ou relação entre type e token (que indica a riqueza lexical do texto – quanto

maior a relação, mais variado é o léxico do texto) e a *standardised type-token*, usada para minimizar o impacto do tamanho do texto na *type-token ratio*.

O cálculo do número de sentenças e de parágrafos é feito com base em convenções de pontuação e será errôneo se os textos utilizados não seguirem tais convenções.

A segunda ferramenta, *KeyWords*, é usada para gerar listas de palavras-chave a partir da lista gerada pela *WordList*.

Para gerar as listas de palavras-chave, além da lista de palavras do corpus gerada pela *WordList*, é preciso ter um corpus de referência. O corpus de referência deve ser de língua geral e preferencialmente cinco vezes maior que o corpus de estudo para garantir que as palavras-chave sejam realmente especificidades do corpus de estudo. As palavras-chave são aquelas que ocorrem com mais frequência, estatisticamente, no corpus de estudo do que no corpus de referência.

Vale observar que a coleta de um corpus de referência é normalmente ainda mais trabalhosa e criteriosa do que a do corpus de estudo. O ideal é utilizar corpora de referência já disponíveis para venda ou download gratuito.

A ferramenta *KeyWords* também pode gerar a lista de palavras-chave-chave, que são as palavras-chave comuns a mais de um texto do corpus de estudo. Essa é possivelmente a lista mais relevante no contexto de elaboração de glossários, já que ela apresenta as palavras-chave que mais ocorreram em todo o corpus.

A terceira ferramenta é o *Concord*, que gera “concordâncias ou listagens de ocorrências de um item específico (chamado palavra de busca ou nóculo, que pode ser formado por uma ou mais palavras) acompanhado do texto ao seu redor (o contexto)” (Sardinha, 2004:104). Essa ferramenta é especialmente útil para análise de termos compostos e contexto de uso, sendo assim também relevante para elaboração de glossários bilíngües de especialidade em que se deseje incluir os contextos para os termos.

Para especificar um determinado item que deve concorrer com o nóculo (por exemplo, *casa* e *sogra*), basta digitá-lo na caixa *Context Word*. A ferramenta também possibilita a especificação do horizonte (número de palavras à direita e à esquerda do nóculo) e a direção (somente à direita ou somente à esquerda, ou ainda X itens à direita e Y itens à esquerda).

Também é possível analisar os *colocados*, que são as palavras que concorrem com o nóculo, e seus contextos.

Considerações finais

Este artigo apresentou o fenômeno da globalização que acelera o desenvolvimento desta disciplina, apoiando-se na criação de novas técnicas e materiais de apoio práticos. Segundo (Krieger e Finatto, 2004), a terminologia, hoje em dia, é concebida como uma disciplina autônoma e multidisciplinar, que demanda conhecimentos lingüísticos, cognitivos, lógicos, ontológicos e da ciência de informação. A quantidade de informações que pode ser processada e analisada e a rapidez com que isso pode ser feito possibilitam observações e estudos que seriam impossíveis manualmente, além de conferirem um caráter de imparcialidade e representatividade à pesquisa.

Para que isso aconteça de forma adequada, no entanto, é imprescindível que a coleta e a criação de corpus sejam feitas segundo critérios rigorosos que, não sendo respeitados, podem comprometer a pesquisa como um todo, induzindo a resultados equivocados.

Atualmente, é possível que a utilização de ferramentas como o *Wordsmith Tools* ainda apresente algumas limitações para os tradutores individualmente, devido à complexidade de elaboração do corpus, aos recortes de conteúdo do material fornecido pelo cliente para tradução (não necessariamente abrangente sobre um determinado tema), aos custos de aquisição de corpus de referência e da própria ferramenta e ao tempo normalmente escasso de que os tradutores dispõem para o trabalho. No entanto, para empresas de tradução, a ferramenta pode vir a ser útil para a geração de glossários e guias de estilo. Diferentemente dos tradutores, essas empresas dispõem de pessoal e recursos dedicados à terminologia e têm acesso a um vasto material de um mesmo cliente devido ao volume de palavras dos projetos que realizam. Para essas empresas, tais ferramentas podem apresentar um custo acessível e capaz de apresentar os dados de forma mais abrangente, com contexto e mais imparcialidade. Mesmo para os tradutores individualmente, a familiaridade com essa e outras ferramentas de corpus torna-se imprescindível.

Como aponta Bowker “a tradução é uma profissão que exige aprendizado permanente, o que significa que os tradutores profissionais estão sempre adquirindo

novas habilidades” (2004:239, minha tradução). O perfil do mercado pode mudar, nossos clientes podem mudar, e ter essas experiências amplia a gama de possibilidades de atrair clientes. Os preços dessas ferramentas devem diminuir a entrada no mercado de profissionais que receberam treinamento para uso dessas ferramentas na universidade deve aumentar. As necessidades de cada cliente, de cada projeto, são diferentes e é importante conhecer as ferramentas para poder escolher.

Referências

- BOWKER, Lynne (2004). “Corpus resources for translators: academic luxury or professional necessity?”. Em: *TradTerm*, volume 10. São Paulo: Humanitas FFLCH/USP.
- FINATTO, Maria José Bocorny. *Introdução à Terminologia: teoria e prática*. São Paulo: Contexto, 2004.
- KRIEGER, Maria da Graça; Bevilacqua, Cleci. A pesquisa terminológica no Brasil: uma contribuição para a consolidação da área. *Revista debate terminológico*, n. 01, 03/ 2005, 2004. Disponível em: www.riterm.net/revista/n_1/kriger.pdf
- SARDINHA, Tony Berber. *Linguística de Corpus*. São Paulo: Manole, 2004.
- TANGNIN, Stela. *O jeito que agente diz. Expressões convencionais e idiomáticas*. São Paulo: Disal Editora, 2005.
- _____. (2004). “Corpora: o que são e para quê servem”. On line: <http://www.fflch.usp.br/dlm/comet/>
- TEIXEIRA, Elisa Duarte. Tese de doutorado. *A linguística de corpus a serviço do tradutor: proposta de um dicionário de culinária voltado para a produção textual*. São Paulo, 2008.
- TradTerm: Revista do Centro Interdepartamental de Tradução e Terminologia/ Faculdade de Filosofia, Letras e Ciências Humanas. *Linguística de Corpus e Tradução técnica- Relato da montagem de um corpus multivarietal de culinária*. Universidade de São Paulo, 1994.
- D.E.L.T.A. Vol. 16 N 2, 2000. *Linguística de Corpus: Histórico e problemática*. Tony Berber Sardinha.