

Elaboração de um Glossário Trilíngue para a área de Tecnologia de Alimentos

Victor Ernesto Silveira da Silva¹

Resumo

Similarmente a diversas áreas científicas e tecnológicas, no campo de Tecnologia de Alimentos, os resultados de pesquisas, produtos novos e as mais recentes descobertas são, muitas vezes, veiculados em línguas estrangeiras, sobretudo em Inglês e Espanhol. Comumente, os profissionais aguardam que outros façam as devidas traduções para que tenham acesso a tais informações. A tarefa da tradução, contudo, é um trabalho complexo e o texto traduzido para português pode ser lançado tardiamente. Ademais, percebe-se a escassa quantidade de glossários e dicionários especializados que facilite a leitura de materiais veiculados em inglês e espanhol para profissionais e estudantes da área de Tecnologia de Alimentos dificultando significativamente o acesso às inovações internacionais nas áreas relacionadas a Tecnologia de Alimentos. Levando-se em consideração o exposto, este projeto de pesquisa objetivou: investigar e listar os termos técnicos mais comuns da área de Tecnologia de Alimentos e traduzi-los para Inglês e Espanhol, a fim de produzir e publicar um glossário trilíngue. Para tanto foram adotados os aportes teóricos e a metodologia da Linguística de Corpus, cuja ideia principal é compilar as palavras mais frequentes de textos de uma determinada área para a investigação de comportamentos linguísticos, frequência linguística e estudos linguísticos sincrônicos e diacrônicos. A Linguística de Corpus inclui, obrigatoriamente, a utilização de programas específicos de computadores que são capazes de fazer a listagem, contagem e verificação de frequência de palavras em textos. Neste projeto adotou-se o programa AntConc 3.4.4 desenvolvido pela Universidade de Waseda no Japão e disponibilizado gratuitamente pelo desenvolvedor. Por meio deste programa e, a partir de 200 textos científicos da área de Tecnologia de Alimentos foram listados 1134 termos frequentes e relevantes para o campo, os quais foram traduzidos em Inglês e Espanhol e publicados em formato de livreto.

Palavras-Chave: Glossário. Alimentos. Trilíngue.

Abstract

Similarly to several scientific and technological areas, in the field of Food Technology, the results of researches, new products and innovations are often conveyed in foreign languages (FL) suchs as Spanish and English. In Brazil, professionals of those areas have to wait for a Portuguese translation to have access to the information publicized in FL. Translation is a long and hard task, occasionally thee work translated may be outdated if compared to the velocity of the science developments. Moreover the sparse number of specialized glossaries and dictionaries makes it difficult to Food Technology researchers and students to translate FL materials for their own. Considering the context above, this article comprises the discussion about the development of a Food Technology Trilingual Glossary for students and researchers in the Federal Institute Baiano campus Catu Bahia-Brazil. The project aimed to investigate, list and translate from Portuguese to English and Spanish the most common technical words in Food Technology and gather them in a Trilingual Glossary. The theory and methodology used as base for the work was Corpus Linguistics (CL). This methodology tends to list the most frequent words in determined texts in order to investigate linguistic behaviors, linguistic frequency, and diachronical and synchronical investigations on language. It is mandatory in CL the use of computer progammes for the data processing. The project described here adopted the programm AntConc 3.4.4 developed by Lawrence Anthony at Waseda University – Japan. The programm is freely avaiable on the Internet. As from more then 200 texts on Food Technology and using AntConc 3.4.4, it was listed more than 500 frequent words in the field. The words were considered very relevant for the theme of the Glossary. Those entries were translated to English and Portuguese and listed. The work was printed as booklet and distributed to Federal Institutes.

Keywords: Glossary. Food. Trilingual.

¹ Mestre em Estudos de Linguagem. Professor do Instituto Federal de Educação, Ciência e Tecnologia da Bahia.

1 Introdução

Este artigo tem como objetivo descrever o processo de produção de um Glossário Trilíngue da área de Tecnologia de Alimentos. O curso em questão tem como um dos objetivos, a formação de profissionais para uma área que basicamente atende a grandes indústrias que se valem de pesquisas e regulamentações, que muitas vezes, são veiculadas em língua estrangeira, portanto, é de suma importância que os estudantes sejam incentivados e expostos à produção/recepção de textos orais e escritos, em inglês e espanhol. De acordo com os professores da área de Tecnologia de Alimentos atuantes no Instituto Federal Baiano campus Catu, os textos que trazem inovações tecnológicas, em sua maioria, são publicados em línguas estrangeiras e, geralmente, os estudantes não possuem proficiência linguística para lê-los ou, até mesmo, recursos para traduzi-los para a língua materna. Essa informação é corroborada pelo fato da biblioteca do campus não contar com material didático ou consultivo (como dicionários) que favoreçam o ensino/aprendizagem da língua estrangeira com foco na área específica do curso. Deste modo, torna-se evidente a necessidade da elaboração de um glossário trilíngue e sua divulgação para a comunidade acadêmica. Assim sendo, o projeto procurou suprir de forma pontual a carência de material didático e consultivo para os estudantes de cursos profissionalizantes e, possivelmente, para pesquisadores, engenheiros, nutricionistas e tecnólogos de áreas afins a Tecnologia de Alimentos.

De modo geral, o projeto que serviu de fundamento para a produção do glossário trilíngue, tinha como objetivo: investigar e listar os termos técnicos mais comuns da área de Tecnologia de Alimentos e traduzi-los para Inglês e Espanhol, com vistas à produção e publicação de um glossário trilíngue nos formatos impresso e virtual.

A metodologia e fundamentação teórica foram pautadas na Linguística de Corpus (LC), sendo esta, a primeira etapa do projeto: a revisão bibliográfica fundamentada na Linguística de Corpus mais especificamente na elaboração de glossários.

Partindo dos princípios norteadores da LC, o projeto percorreu o seguinte trajeto:

- Reunião e análise dos textos autênticos em língua portuguesa na área Tecnologia de Alimentos, seguindo as indicações bibliográficas das professoras de Engenharia de Alimentos, Tecnologia de Alimentos e Nutrição do campus.
- Listagem das palavras que aparecem com mais frequência nos textos analisados com a utilização do programa AntConc 3.4.4 desenvolvido por Lawrence Anthony da Universidade de Waseda no Japão;
- Seleção das palavras mais representativas da área com o auxílio de profissionais de Tecnologia de Alimentos e agrupá-las por campo semântico;
- Tradução contextualizada das palavras elencadas para as línguas Inglesa e Espanhola;
- Publicação do glossário nos modos impresso e virtual.

Vale ressaltar que, o projeto Glossário Trilíngue foi fomentado pela Pró-reitoria de Pesquisa do Instituto Federal Baiano, o que conferiu aos pesquisadores e bolsistas a garantia da produção efetiva do Glossário impresso e o reconhecimento desta iniciativa como relevante para a instituição.

2 Linguística de Corpus e a Produção de Glossários

A produção de glossários geralmente é fruto de pesquisas dentro das áreas de Lexicografia e Terminologia que possuem seus métodos e aportes teóricos específicos. Enquanto ciência maior, a Lexicografia trata das palavras de uma língua e seus sentidos no intuito de registrá-las em dicionários e vocabulários. A terminologia, por sua vez, ocupa-se de vocabulários de áreas específicas, geralmente, de natureza técnica. Recentemente, porém, tem crescido estudos de natureza lexicográfica, fundamentados na Linguística de Corpus (LC).

Na produção do Glossário Trilíngue da área de Tecnologia de Alimentos, adotou-se a LC e quais contribuições essa ciência trouxe para o desenvolvimento e avanço da pesquisa linguística empreendida. Segundo Sardinha:

A Linguística de Corpus ocupa-se da coleta e exploração de corpora, ou conjuntos de dados linguísticos textuais que foram coletados criteriosamente com o propósito de servirem para a pesquisa de uma língua ou variedade linguística. Como tal, dedica-se à exploração da linguagem através de evidências empíricas, extraídas por meio de computador. (SARDINHA, 2000, p.2)

Assim, a LC abarca pesquisas que envolvem a elaboração de glossários, análises linguísticas diacrônicas e sincrônicas entre outros, partindo de corpora coletados e analisados por programas específicos de computador. Levando-se em consideração a definição de corpus feita por Sardinha (2000), os textos utilizados precisam ser autênticos, produzidos por pessoas nativas e devem ser escolhidos de forma minuciosa de modo a serem representativos na área que se pretende empreender a análise linguística. Um conceito fundamental para o trabalho em LC é o de representatividade, o qual segundo Sinclair (1991) é proporcional ao tamanho do corpus. Para este trabalho será considerada a classificação de representatividade por extensão de corpora apresentado por Sardinha (2000):

Tamanho em palavras	Classificação
Menos de 80 mil	Pequeno
80 a 250 mil	Pequeno-médio
250 mil a 1 milhão	Médio
1 milhão a 10 milhões	Médio-grande
10 milhões ou mais	Grande

Quadro 1: Classificação de Corpus por Sardinha (2000)
Fonte: SARDINHA (2000)

Sendo assim, optou-se por trabalhar um corpus de pequeno porte, pois a elaboração de um glossário trilingue voltado para uma área específica é mais um trabalho terminológico do que lexicográfico, ou seja, concluiu-se posteriormente que, por ser uma área multidisciplinar, Tecnologia de Alimentos contava com um número reduzido de termos específicos da área. Vale ressaltar que o caráter específico da LC é a utilização de corpus eletrônico e a compilação de dados feita com programas específicos de criação de listas de palavras. Os programas de compilação de palavras a partir de corpus eletrônicos mais conhecidos são *WordSmith Tools* e *AntConc*, que além de listar palavras, destacam a frequência das mesmas e assinalam outros efeitos como: concordância, palavras chaves e referência a corpus maiores existentes em grandes centros universitários como a Universidade de Oxford na Inglaterra (ALUISIO;ALMEIDA,2006).

Por muito tempo, apenas grandes centros de estudos linguísticos detinham a tecnologia de compilação de palavras em textos, mas com o avanço e popularização da

Internet, vários programas foram desenvolvidos e disponibilizados na rede, no entanto, nem todos são gratuitos.

3 Construção de um Glossário: metodologia

Para dar conta da tarefa de listagem de palavras foi necessária a adoção de uma metodologia que abarcasse a tarefa de compilação, análise e tradução de palavras de uma determinada área de conhecimento. Portanto, foi adotada a Linguística de Corpus, que é uma ciência dotada de uma metodologia específica para a análise do léxico em um ou mais textos dentro de um determinado contexto. Segundo Sardinha (2000) a Linguística de Corpus ocupa-se da coleta e exploração de conjuntos de dados linguísticos textuais que foram coletados para a pesquisa de uma língua. Assim, dedica-se à exploração da linguagem através de evidências empíricas, coletadas por meio de computador.

Percebe-se que a pesquisa teve notadamente um caráter quantitativo, uma vez que a maior parte do trabalho foi o de transformação dos textos em listas e a construção de conjuntos de palavras significativas para a área temática do glossário. Vale ressaltar, no entanto, que, o estudo da relevância dos textos usados na compilação e a categorização das palavras em significativas ou não foi um esforço qualitativo, pautado no conhecimento da área de tecnologia de Alimentos das bolsistas e das orientações das professoras e engenheiras de alimentos. Assim, foram determinadas 7 etapas para a produção do Glossário:

A primeira etapa do trabalho aqui discutido foi um curso básico de LC para as bolsistas, para uma rápida compreensão dos aportes teóricos da LC, bem como aprofundar-se nas teorias sobre elaboração de glossários com tradução de palavras.

A segunda etapa constituiu-se na reunião e análise de textos de relevância na área de Tecnologia de Alimentos fornecidos virtualmente em formato *Open Document Format* (ODF), Documento do Microsoft Word (DOC) ou *Portable Document Format* (PDF) pelas professoras: Karina Pitta e Solane Alves do Instituto Federal Baiano campus Catu. Tal escolha foi feita tendo como base o conhecimento técnico que tais professoras possuem nas áreas de: Engenharia de Alimentos, Nutrição, Gastronomia, Bromatologia entre outras áreas que integram o curso de Tecnologia em Alimentos. Foram reunidos 37 artigos científicos em

língua portuguesa produzidos entre os anos 2000 e 2015 e publicados em periódicos ou livros online.

Na terceira etapa foi feita a conversão dos textos nos formatos ODF, DOC e PDF para o formato TXT, uma vez que o programa de compilação de palavras leem arquivos somente neste molde;

Na quarta etapa ocorreu a listagem das palavras por meio do programa AntConc 3.4.4 desenvolvido por Laurence Anthony da Universidade de Waseda no Japão, o qual elabora listas de palavras mais frequentes em um determinado texto como pode ser visto na figura 1.

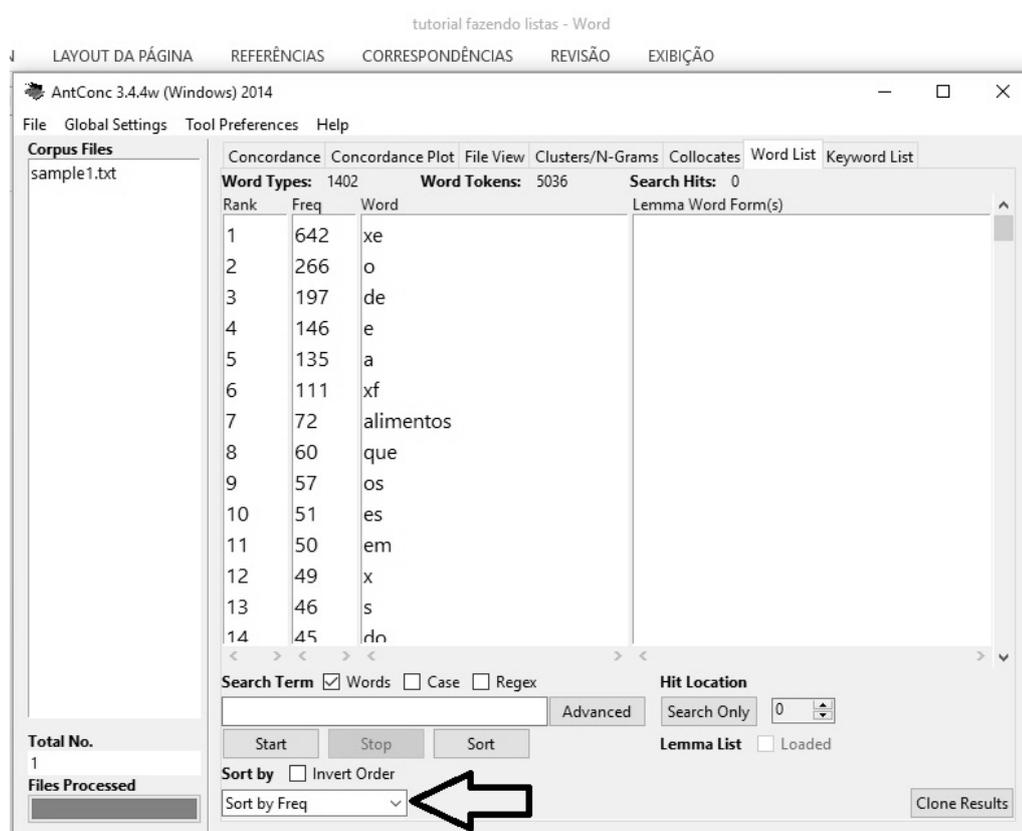


Figura 1. Processo de construção de listas com o Ant.Conc 3.4.4
Fonte: o autor

Nota-se na figura acima o *Rank* da palavra determinada pela sua Frequência (representada pela abreviação *freq*), ou seja no primeiro lugar esta a sílaba *xe* que ocorreu 642 vezes no texto. O problema é que *xe* não é uma palavra na língua portuguesa, mas provavelmente é uma sílaba que se formou no processo de conversão entre os arquivos (de PDF para TXT, por exemplo) uma vez que muitas vezes ocorrem inconsistências por parte do software de conversão de textos. Entre os ranks 2 e 5 tem-se palavras extremamente comuns na língua portuguesa. Os artigos **o** e **a**, a preposição **de** e o pronome relativo **que** são palavras gramaticais essenciais para a língua portuguesa pois sempre aparecem em discursos sejam eles orais ou escritos. Era de se esperar que a frequência destas palavras fosse alta. No rank 7 aparece a primeira palavra relevante para o trabalho: **alimentos** que aparece 72 vezes em um determinado texto. Esta tarefa foi feita em cada um dos 200 textos selecionados e assim foram listadas 1005 palavras. Na quinta etapa foi feita a limpeza de texto, ou seja, a eliminação de palavras não significativas para os propósitos da listagem. A eliminação de palavras foi feita segundo os seguintes critérios:

- a) Palavras gramaticais: foram eliminadas da listagem preposições, conjunções, verbos flexionados, pronomes e interjeições. Os substantivos próprios também foram eliminados. Por fim, o glossário contou apenas com: substantivos comuns, adjetivos, verbos no infinitivo e advérbios.
- b) Palavras comuns: foram consideradas palavras comuns, as palavras que pertenceriam comumente a qualquer ciência por exemplo: “livro”, “computador” ou “laboratório”. Possivelmente uma ou outra palavra comum foi mantida devida a sua relação essencial com palavras relevantes para a área de Tecnologia de Alimentos como por exemplo: “fruta”, “alimento”, “sal” e “procedimento”.
- c) Palavras de baixa frequência: o AntConc 3.4.4 agrupa as palavras e conta quantas vezes a palavra aparece nos textos. Essa contagem foi um dos critérios para determinar a relevância da palavra para a área temática do glossário. Palavras que, geralmente circulam no discurso científico e que possuíam a frequência igual a 1 em um texto bem como não recorrência desta palavra em outros textos levou a eliminação da mesma, a exemplo disso tem-se

“formol”, “vidraria” e “sistema digestório”. Em seguida, as palavras foram agrupadas em ordem alfabética e classificadas quanto a classe gramatical. O quadro 2 abaixo ilustra a organização

1	Abacaxi	sub.
2	Abacate	sub.
3	abastecimento	sub.
4	absorção	sub.
5	acarretar	vb.
6	aceitação	sub.
7	aceleração	sub.
8	acentuado	adj.
9	Acerola	sub.
10	acetato	sub.
11	acidez	sub.
12	Acidificação	sub.
13	Acido cítrico	sub.
14	Acido fólico	sub.
15	Acidulante	sub.

sistemizada para o

glossário:

Quadro 2. Exemplo de organização do glossário trilingue

Fonte: os autores

Pode-se notar que algumas palavras como “absorção” e “aceitação” seriam facilmente categorizadas como palavras comuns, no entanto, elas são importantes para compreensão de processos fundamentais na área de tecnologia de Alimentos e portanto merecem o destaque e a devida tradução. Ademais, estas palavras foram muito frequentes nos 200 textos coletados para a formação das listas de palavras. O suporte técnico dado pelas professoras e engenheiras de alimentos do Instituto Federal Baiano campus Catu também consolidou a importância das palavras citadas.

A sexta etapa foi marcada pela tradução das palavras para o Inglês e Espanhol com a ajuda de dicionários gerais impressos e virtuais, bem como, a consulta aos profissionais da área de ensino de línguas estrangeiras, o quadro 3 demonstra a organização para a tradução das palavras selecionadas:

1	Abacaxi	sub.	<u>pineapple</u>	<u>Piña</u>
2	Abacate	sub.	<u>Avocado</u>	<u>palta, Aguacate</u>
3	abastecimento	sub.	<u>supply</u>	<u>Aprovisionamiento</u>
4	absorção	sub.	<u>absorption</u>	<u>absorción</u>
5	acarretar	vb.	<u>To entail</u>	<u>ocasionar</u>
6	aceitação	sub.	<u>acceptance</u>	<u>aceptación</u>
7	aceleração	sub.	<u>acceleration</u>	<u>Aceleración, aceleramiento</u>
8	acentuado	adj.	<u>accented</u>	<u>destacado</u>
9	Acerola	sub.	<u>Barbados cherry</u>	<u>acerola</u>
10	acetato	sub.	<u>acetate</u>	<u>acetato</u>
11	acidez	sub.	<u>acidity</u>	<u>acidez</u>
12	Acidificação	sub.	<u>Acidification</u>	<u>acidificación</u>
13	Acido cítrico	sub.	<u>Citric acid</u>	<u>acido citrico</u>
14	Acido fólico	sub.	<u>Folic acid</u>	<u>acido fólico</u>
15	Acidulante	sub.	<u>Acidulant</u>	<u>Acidulante</u>
16	acondicionado	adj.	<u>conditioning</u>	<u>Acondicionado -</u>
17	açucar	sub.	<u>sugar</u>	<u>azucar</u>
18	adequado	adj.	<u>suitable</u>	<u>Apropiado</u>
19	adição	sub.	<u>add</u>	<u>adición</u>
20	Aditivos	sub.	<u>Additives</u>	<u>aditivo</u>
21	Aeróbico	adj.	<u>Aerobic</u>	<u>aerobic</u>
22	aeração	sub.	<u>aeration</u>	<u>aieración</u>
23	Agitador magnético	sub.	<u>Magnetic stirrer</u>	<u>agitador magnético</u>
24	agricultor	sub.	<u>farmer</u>	<u>agricultor</u>
25	agricultura	sub.	<u>Husbandry, agriculture</u>	<u>agricultura</u>
26	agridoce	adj.	<u>bittersweet</u>	<u>Agridulce</u>
27	agroindustrial	adj.	<u>agroindustrial</u>	<u>agroindustrial</u>

Quadro 3. Trabalho de tradução das palavras selecionadas

Fonte: os autores

Na última etapa, foi feita a revisão final do material com a participação dos profissionais da área de tecnologia de alimentos, em seguida o material foi impresso no formato Livroto colorido (figura 2.).



Figura 2. Produto final do Projeto Glossário Trilíngue na área de Tecnologia de Alimentos

Fonte: os autores

Na organização do material a ser impresso optou-se por separar o glossário em três sessões: Português-Espanhol, Intervalo e Português-Inglês. Inicialmente a ideia era preencher o intervalo com um vocabulário de comidas típicas brasileiras traduzidas para o inglês e o espanhol. No entanto, para a manutenção da fidelidade com o conteúdo das listas compiladas dos textos, decidiu-se tornar o Intervalo em um vocabulário de frutas. Isto porque os pratos regionais brasileiros seriam inseridos voluntariamente pelas bolsistas, visto que, poucas comidas regionais foram mencionadas nos 37 textos-base para a construção do glossário. As frutas, contudo, apareceram com frequência. Muitas delas não possuem tradução para língua inglesa ou espanhola como “pequi” ou “jabuticaba”, neste caso, foram utilizadas as imagens da fruta incluindo o respectivo nome científico.

O resultado do projeto de pesquisa aqui descrito, foi a elaboração do glossário trilingue para uso dos alunos da área de Tecnologia de Alimentos da comunidade acadêmica do Instituto federal Baiano, contando com 1134 palavras traduzidas para Inglês e Espanhol. Neste total foram identificados 991 substantivos, 109 adjetivos, 30 verbos e 4 advérbios. A escolha dos verbos e advérbios a serem mantidos na lista do glossário se deu com base na indicação das profissionais da área de Tecnologia de Alimentos que avaliaram a relevância

dos termos técnicos. A mesma consultoria foi prestada ao se decidir na manutenção de palavras compostas como “ácido fólico” ou “agitador magnético”. Isto porque o AntConc lista os termos apenas como palavras simples, de modo que os substantivos compostos têm suas palavras separadas no processo de listagem. Decidiu-se, portanto, identificar as palavras que, nos textos, fazem parte de substantivos compostos e mantê-las enquanto palavras compostas visto que separadas, as palavras não representavam relevância para a área de Tecnologia de Alimentos. A exemplo disso tem-se “rigor mortis” listadas separadamente como “rigor” e “mortis”. Note-se que ambas palavras são dispensáveis para os propósitos do glossário. Foi necessária a supervisão técnica das professoras Karina Pitta e Solane Alves para a decisão da manutenção ou não destes termos como compostos ou simples. Deste modo, foram listadas 13 palavras compostas e 1121 palavras simples.

4 Considerações Finais

Com o projeto de pesquisa aqui descrito, procurou-se associar o ensino e aprendizagem de línguas estrangeiras com a pesquisa voltada para a área técnica, possibilitando, assim, aos professores do curso de tecnologia de Alimentos a utilização de textos em inglês e espanhol em suas aulas. Concomitantemente, os professores de língua estrangeira podem fazer uso destes mesmos textos técnicos com objetivo de estudos linguísticos. Por fim, o produto desta pesquisa tem caráter permanente e pode ser utilizado por diversas gerações de alunos e professores com diferentes finalidades, desta forma percebe-se a articulação entre ensino, pesquisa e extensão.

É sempre válido ressaltar a importância do trabalho aqui descrito. Em um estudo piloto foram feitas buscas no meio virtual e percebeu-se o número limitado de glossários multilíngues voltados para a área de Tecnologia de Alimentos. No caso de glossários trilíngues mais especificamente em Português, Inglês e Espanhol não foram encontrados resultados significativos. Na biblioteca do Instituto Federal Baiano campus Catu, constatou-se a inexistência de exemplares que possibilitem aos alunos a consulta de termos em Inglês e Espanhol que façam parte de seu repertório técnico. A própria confirmação da inexistência destes glossários corrobora o caráter inovador deste projeto. Contudo, esta inovação precisa

ser estendida para outras áreas, tais como Química, Agropecuária, Petróleo ou Agrimensura (para citar as áreas de formação de técnicos no Instituto Federal Baiano campus Catu). Existem inúmeros motivos pelos quais os profissionais e cientistas da atualidade necessitam aprender e utilizar uma segunda língua e é função da escola aproximar o estudante deste universo e mostrar os benefícios do multilinguagem.

É relevante notar também que o glossário produzido poderia ser maior, utilizando um corpus maior e mais abrangente. Para tanto seriam necessários mais textos da área de tecnologia de alimentos e mais tempo para a compilação das palavras. Ademais, o grupo de pesquisa deveria ser mais numeroso para dar conta do volume de palavras a serem analisadas e traduzidas. Essa ideia é bastante promissora, uma vez que poderia trazer para o glossário o maior número possível de palavras independentemente de sua frequência nos textos. De qualquer sorte, como material consultivo para o ensino integrado técnico e profissionalizante, este glossário será um grande facilitador para a compreensão e produção de textos em língua espanhola e inglesa.

Referências

ALÚSIO, Sandra Maria; ALMEIDA Gladis Maria de Barcellos. O que é e como se constrói um *corpus*? Lições aprendidas na compilação de vários *corpora* para pesquisa lingüística. *Calidoscópio*. V. 4, n. 3, p. 156-178, set/dez 2006

ANTHONY, L. *Laurence Anthony's website*. *AntConc* 3.4.4. Disponível em <<http://www.laurenceanthony.net/software.html>> Acesso em Ago. 2015

SARDINHA, Tony Berber. *Lingüística de Corpus: Histórico e Problemática*

DELTA. V.16, n.2 São Paulo, 2000

SINCLAIR, J. *Corpus, concordance, collocation*. Oxford: Oxford University Press. (19991)

Data de submissão: 17/04/2017. Data de aprovação: 06/06/2017