

A arquitetura de um glossário terminológico Inglês-Português na área de Eletrotécnica

The architecture of an English-Portuguese glossary in the field of Electrical Terminology

Sabrina Bonqueves Fadanelli
Universidade Federal do Rio Grande do Sul (UFRGS)
sabrina_mina2006@hotmail.com

Maria José Bocorny Finatto
Universidade Federal do Rio Grande do Sul (UFRGS)
mariafinatto@gmail.com

Resumo

Neste artigo descrevemos alguns dos procedimentos para a execução de um protótipo online de glossário Inglês-Português na área de Eletrotécnica/Engenharia Elétrica – voltado principalmente aos alunos iniciantes dos cursos técnicos de Eletrotécnica e de graduação em Engenharia Elétrica/Eletrotécnica. A metodologia envolveu um *corpus* de *datasheets*, documentos muito utilizados por profissionais das áreas de Elétrica e Eletrotécnica; e a comparação com os dados obtidos de 108 alunos iniciantes das áreas Elétricas. Os resultados apontam para a relevância de se considerar o olhar do público-alvo para a compilação do glossário.

Palavras chave

Terminografia, glossário, *datasheet*, Eletrotécnica

Abstract

This article describes some of the procedures for the execution of an online English-Portuguese glossary prototype in Electrical Engineering / Electrotechnical Field terminology – aimed mainly at beginner students from technical and graduation courses in Electrical Engineering. The methodology is comprised of a corpus of *datasheets*, documents often used by professionals of the Electrical Engineering area, and the comparison of data obtained from these *datasheets* with the data gathered from 108 students of Electrical courses. Results point to the relevance of considering the point of view of our target audience to build the glossary properly.

Keywords

Terminography, pedagogical, glossary, *datasheet*, electrotechnical field

1 Introdução

Até que se tornem especialistas ou proficientes na terminologia de sua área, estudantes de Eletrotécnica e Engenharia Elétrica no Brasil devem aprimorar seus conhecimentos em Língua Inglesa, já que praticamente toda a documentação de componentes elétricos se encontra nesta língua. A Eletrotécnica é uma área da Engenharia Elétrica que lida com instalações de redes elétricas, transformadores, circuitos elétricos residenciais, industriais, etc. Seu material principal de consulta são os *datasheets*, documentos que contêm em sua estrutura informações técnicas sobre os dispositivos e aparelhos elétricos. Muitos dos estudantes iniciantes nestas áreas encontram dificuldades na leitura destes documentos técnicos, já que muitas vezes não possuem um conhecimento proficiente da Língua Inglesa, nem da área técnica em si. Isso demonstra uma necessidade de se produzir uma ferramenta que forneça tanto a tradução dos termos de Eletrotécnica em português, como também apresente definições explicativas e didáticas voltadas aos alunos iniciantes da Eletrotécnica/Engenharia Elétrica ou a possíveis usuários que apresentem necessidade de utilização da ferramenta. Além disso, a ferramenta deverá conter exercícios que possam auxiliar na leitura e compreensão dos *datasheets*. O desenvolvimento desta ferramenta, assim sendo, deverá seguir parâmetros pertencentes a uma terminografia que atenda às necessidades deste público-alvo.

O objetivo deste artigo é descrever alguns procedimentos usados para a seleção de termos que venham a integrar um protótipo de glossário online Inglês-Português na área de Eletrotécnica/Engenharia Elétrica, sob a ótica da

Terminologia Textual.

A perspectiva da Terminologia Textual considera o texto especializado como o habitat natural das terminologias, sendo ele o todo que vai determinar seus modos de dizer específicos (Kilian, 2007; Zilio, 2010). Com base nisso, não só a terminologia em si é focada, mas também outros aspectos textuais como macro e microestruturas do texto: as frases, os sintagmas, os gêneros textuais, etc. (Finatto, 2004, 2010). Como afirmam Bourugault & Slodzian (2004, pg. 107), a abordagem textual permite visualizar todas as “experiências” da análise linguística textual, não considerando os termos como “unidades de conhecimento que habitam a língua”.

2 Os datasheets

Os documentos técnicos chamados de *datasheets* fornecem informações como potência, resistência, dimensões, etc., não somente com a finalidade de compreender o funcionamento de um componente, mas também para não danificá-lo e garantir a segurança do usuário (ver Figura 1 para exemplos de *datasheets*). De acordo com Dewey (1998), o *datasheet* é o conector comunicativo entre indivíduos ou empresas com necessidades diferentes e níveis de conhecimento diferentes. A relação entre os engenheiros que escrevem os *datasheets* e os seus possíveis leitores pode variar: comprador do produto especificado pelo *datasheet* x vendedor; equipe de produção e engenharia do produto, pessoal do *marketing*, e inspetores de qualidade, além de outros.

3 A Metodologia

Nesta investigação, escolhemos a metodologia da Linguística de Corpus como apoio. A Linguística de *Corpus* faz uso da compilação e análise de elementos e estruturas linguísticas de um determinado *corpus* por meio de sistemas computadorizados (Berber-Sardinha, 2004; Biber, 1988). A utilização da Linguística de *Corpus* nos permitirá analisar a ocorrência de termos nos *datasheets*.

Além disso, levamos em consideração não somente o material de onde os termos serão retirados, mas também o ponto de vista das pessoas que o utilizarão: os alunos, fazendo uma comparação entre os dados obtidos de ambas as fontes.

A obtenção de dados de extratores foi realizada utilizando duas ferramentas, o AntConc (Anthony, 2004) e o TermoStat (Drouin, 2003). Estas foram escolhidas por serem de natureza diferente, uma estatística e a outra linguística, a

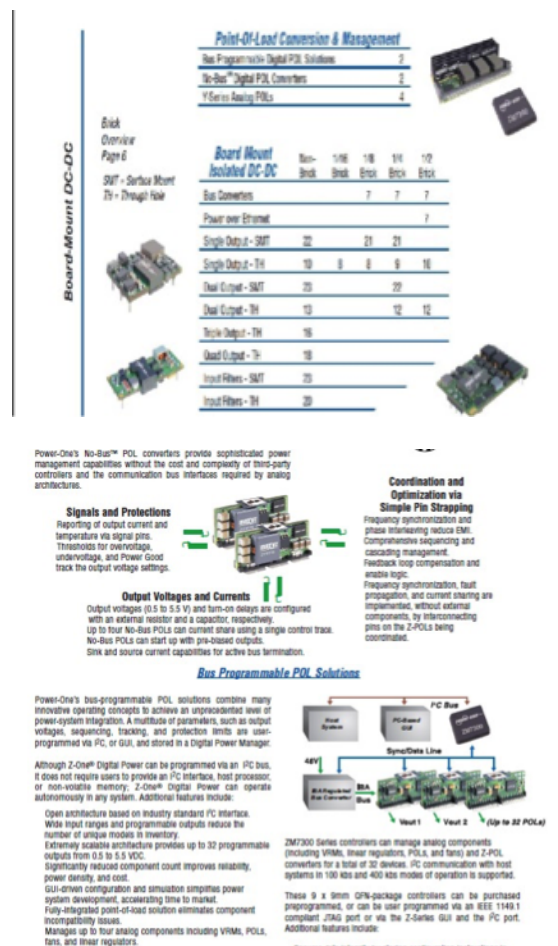


Figura 1: Exemplos de *datasheets*.

fim de promover mais variedade nos resultados (Vieira & Lopes, 2010).

O corpus de estudo foi coletado a partir de *datasheets* em Língua Inglesa de componentes elétricos e áreas da Eletrotécnica sugeridas por dois Engenheiros Elétricos como essenciais para o conhecimento técnico dos aprendizes. Os *datasheets* foram separados em 11 (onze) componentes/áreas pertinentes à Eletrotécnica e transformados em 11 (onze) arquivos em formato txt.

No AntConc, foi utilizada a função *Keyword* — que extrai as palavras-chave de um *corpus*, comparando o *corpus* de estudo com um outro *corpus* — chamado de *corpus* de referência; o *corpus* de referência foi composto de textos extraídos do *corpus Coca*¹ (*Contemporary Corpus of American English*) e do *BNC*² (*British National Corpus*), aleatoriamente. O *TermoStat* (Drouin, 2003) é um extrator de termos acessado gratuitamente na web, que junta candidatos a termos (doravante tratados como CTs) de acordo

¹O *Corpus of Contemporary American English* está disponível em <http://corpus.byu.edu/coca/>.

²O *British National Corpus* está disponível em <http://www.natcorp.ox.ac.uk/>.

com padrões gramaticais e frequência.

Como próximo passo, exatamente os mesmos *datasheets* utilizados na extração automática foram distribuídos a 108 alunos de dois cursos: Técnico em Eletrotécnica da FTEC Faculdades e alunos no primeiro semestre de Engenharia Elétrica da Universidade de Caxias do Sul (UCS). Foi-lhes instruído que marcassem com um círculo ou sublinhassem as palavras dos *datasheets* que eles não compreendiam durante uma leitura cuidadosa do documento. As comparações entre os dados apresentados pelos extratores e pelos alunos levaram em conta as seguintes questões:

- A) Quantos CTs em comum/exclusivos foram apresentados entre os extratores e os dados dos alunos?
- B) Os CTs apontados pelos alunos seriam em sua maior proporção vocabulário pertencente à área de Eletrotécnica/Engenharia Elétrica?
- C) Os CTs que aparecem mais frequentemente nos extratores são os mesmos que os apontados pelos alunos?

3.1 A) Quantos CTs em comum/exclusivos foram apresentados entre os extratores e os dados dos alunos?

Para que o Microsoft Excel separasse os candidatos em comum entre os extratores e alunos, utilizou-se a função *PROCV*. A Tabela 1 permite uma visualização dos totais de termos extraídos do AntConc e TermoStat e apontados pelos alunos. Podemos ver que a quantidade de termos em comum foi baixa.

Total de termos TermoStat	782
Total de termos AntConc	533
Total de termos alunos	640
Termos em comum TermoStat + Alunos	73
Termos em comum AntConc + Alunos	40
Termos em comum AntConc + TermoStat	134

Tabela 1: Termos em comum.

De certa forma este resultado já era esperado, justamente pela natureza diferente entre as duas ferramentas (uma de abordagem estatística e outra de abordagem linguística) e com relação aos alunos, que se utilizam de estratégias cognitivas para a escolha dos termos que desconhecem. A baixa proporção de termos em comum indica a utilidade de se combinar dados dos extratores com dados apontados pelo público-alvo, já que se utilizarmos somente os extratores poderemos não incluir informações importantes para os usuários.

3.2 B) Os CTs apontados pelos alunos seriam em sua maior proporção vocabulário pertencente à área de Eletrotécnica/Engenharia Elétrica?

Um dos questionamentos que se fez presente quando da tentativa de selecionar os termos mais relevantes para os usuários do protótipo de glossário é se os alunos apresentam mais dificuldade em interpretar/compreender termos que pertençam à área de Eletrotécnica, ou, em outras palavras, que sejam conceitos especializados. Desse modo, a fim de se determinar a quantidade de CTs apontados pelos alunos que também possam ser considerados conceitos especializados, decidiu-se por utilizar uma ferramenta online que situe o vocabulário no contexto cultural e socioprofissional da Eletrotécnica/Engenharia Elétrica: a Electropedia³. A Electropedia é um banco de dados terminológicos organizado pela IEC – International Electrotechnical Commission (Comissão Internacional de Eletrotécnica), a organização mundial que prepara e padroniza todas as configurações e características de aparelhos e dispositivos elétricos, voltada em sua essência para especialistas da área, ou ao menos a alguém que possua uma certa fluência em Língua Inglesa.

Manualmente, os CTs em comum entre AntConc, TermoStat e alunos e os CTs exclusivos dos alunos foram inseridos na Electropedia. Não foram inseridos os CTs exclusivos dos extratores pois, o objetivo aqui era verificar o vocabulário apontado pelos alunos. Os CTs que não foram encontrados na Electropedia passaram por uma avaliação de um engenheiro elétrico. A comparação entre os resultados obteve o seguinte: do total de 640 CTs apontados pelos alunos, 295 foram encontrados e 345 não foram encontrados na Electropedia. Destes 345 não encontrados, cerca de 101 eram fórmulas, abreviações e termos que o especialista considerou técnicos, mesmo não aparecendo na Electropedia. Sobraram então 244 CTs que nem o engenheiro considerou como vocabulário técnico, nem foram encontrados na Electropedia. Isso representa cerca de 38% dos CTs apontados pelos alunos.

Embora os dados obtidos respondam à questão de forma a confirmar a predominância de vocabulário pertencente à área de Eletrotécnica/Engenharia Elétrica, os resultados anteriormente descritos mostram que os alunos apresentaram dúvidas em uma porcentagem relevante de CTs não considerados conceitos especializados; conclui-se, então, que incluir estes

³Disponível em Disponível em: <http://www.electropedia.org/>

últimos na compilação do glossário, seja nas definições, seja nos exercícios, provavelmente auxiliaria os usuários em sua leitura dos *datasheets*.

3.3 C) Os CTs que aparecem mais frequentemente nos extratores são os mesmos que os apontados pelos alunos?

Neste estudo, perguntamos se a frequência dos termos já apontados como chave pelos extratores seria um fator que refletiria informações relevantes para a compilação de nosso glossário. Assim sendo, comparamos se os CTs apontados pelos alunos coincidiam em sua maioria com os CTs considerados mais frequentes nos extratores.

Primeiramente fez-se necessário determinar quais CTs nos extratores eram considerados frequentes. Com base na normatização proposta por Biber et al. (1998) a fim de estabelecer aproximadamente quantas vezes uma palavra apareceria em cada mil palavras de um corpus, representando ao menos 1% da frequência (ou dez ocorrências em cada mil), o seguinte cálculo foi usado: quantidade de vezes que um termo aparece, dividido pelo número de termos totais, multiplicado por mil. Todos os 533 CTs resultantes do AntConc já representavam cerca de 1% de frequência, pois a função *keyness* apenas aponta termos que aparecem seis vezes ou mais ($6 \div 533 \times 1000 = +$ ou $- 11$). No TermoStat, determinou-se que 1% de frequência significava 8 ocorrências ($8 \div 782 \times 1000 = 10.23$).

Os CTS foram colocados na planilha *Excel* e novamente a fórmula *PROCV* foi aplicada. A Tabela 2 mostra quantos CTs apontados pelos alunos estão entre os CTs dos extratores considerados mais frequentes.

	AntConc	TermoStat
Total	533	782
Total CTs 1% ou + de frequência	533	64
Total CTs alunos entre os + frequentes dos extratores	142	14

Tabela 2: Comparação de termos frequentes.

Podemos ver que considerar a frequência como um fator determinante na escolha de CTs para o glossário não é muito recomendável. A maioria dos termos que causaram mais dificuldade aos alunos não estão entre os listados como os mais frequentes pelos extratores. Assim sendo, se utilizássemos somente os dados fornecidos pelos extratores estaríamos excluindo dados importantes para a confecção do glossário visando a necessidade dos alunos.

4 Outras Observações

Uma análise mais cuidadosa com os CTs apontados pelos alunos pareceu mostrar que a habilidade de considerar itens periféricos do texto (como combinações de palavras, elementos próximos a figuras, gráficos, reconhecimento de cognatos, etc.) para interpretar um termo aparentemente desconhecido não se fez presente em alguns termos apontados pelos alunos. A análise foi feita examinando cada um dos 640 CTs resultantes da coleta de dados com os aprendizes de Eletrotécnica/Engenharia Elétrica, em seu contexto de ocorrência, com a ferramenta AntConc. Dentre estes cerca de 84 CTs eram cognatos, e muitos dos que não eram cognatos poderiam ter sido compreendidos se levados em conta juntamente com as palavras ou outros elementos que os acompanhavam. Essa característica fornece mais um possível critério para a compilação do glossário: os termos, quando transformados em verbetes no glossário, não devem estar sozinhos — ou seja, cada verbete deve vir acompanhado de seus colocados como apareceram no corpus. E os exercícios devem levar em conta cognatos e as combinatórias lexicais que acompanham cada CT.

Outra questão de grande relevância na análise foram os termos encontrados em tabelas, gráficos e figuras, elementos muito presentes no gênero textual *datasheet*. Cerca de 46% dos termos contidos nestes elementos coincidem com os termos que foram apontados pelos alunos, uma quantidade bastante considerável que indica a importância de serem incluídos no glossário.

4.1 Conclusões

Este estudo nos forneceu dados que permitem afirmar que o desenvolvimento do glossário a partir da consideração do ponto de vista do público-alvo aumenta as chances da ferramenta ser considerada útil por aprendizes ainda não especialistas, pois resulta em uma seleção de termos que vai ao encontro não somente de conceitos especializados que os alunos precisam para ler os *datasheets*, mas também às outras palavras que compõem a tessitura do documento e que possam ser desconhecidas para os alunos.

Referências

Anthony, Laurence. 2004. AntConc: A learner and classroom friendly, multi-platform corpus analysis toolkit. Em *Proceedings of IWLeL: An*

- interactive workshop on language e-learning*, 7–13.
- Berber-Sardinha, Tony. 2004. *Linguística de corpus*. São Paulo: Manole.
- Biber, Douglas. 1988. *Variation across speech and writing*. Cambridge: Cambridge University Press.
- Biber, Douglas, Susan Conrad & Randi Reppen. 1998. *Corpus linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press.
- Bourugault, Didier & Monique Slodzian. 2004. Por uma terminologia textual. Em Maria da Graça Krieger & Luzia Araújo (eds.), *A terminologia em foco. cadernos de tradução* 17, Porto Alegre: Instituto de Letras da UFRGS.
- Dewey, F. Raymond. 1998. A complete guide to datasheets. *Sensors Magazine* Disponível em <http://www.allegromicro.com/~media/Files/Technical-Documents/pub26000-Complete-Guide-To-Datasheets.ashx>.
- Drouin, Patrick. 2003. Term extraction using non-technical corpora as a point of leverage. *Terminology* 9(1). 99–117.
- Finatto, Maria José Bocorny. 2004. Termos, textos e textos com termos: novos enfoques dos estudos terminológicos de perspectiva linguística. Em Aparecida Negri Isquerdo & Maria Graça Krieger (eds.), *As ciências do léxico: lexicologia, lexicografia, terminologia*, vol. 2, 353–363. Campo Grande: UFMS/PPG-Letras UFRGS.
- Finatto, Maria José Bocorny. 2010. Estudos sobre linguagens e textos científicos e técnicos: o que é uma terminologia textual? Em Gisela Collischon (ed.), *Encontro do CELSUL*, 153–172. UNISUL.
- Kilian, Cristiane. 2007. *A retomada de unidades de significação especializada em textos em língua alemã e portuguesa sobre gestão de resíduos: uma contribuição para a tradução técnico científica*. Porto Alegre: Universidade Federal do Rio Grande do Sul. Tese de Doutorado.
- Vieira, Renata & Lucelene Lopes. 2010. Processamento de linguagem natural e o tratamento computacional de linguagens científicas. Em Cristina Becker Lopes Perna, Heloísa Orsi Koch Delgado & Maria José Bocorny Finatto (eds.), *Linguagens especializadas em corpora: modos de dizer e interfaces de pesquisa*, Porto Alegre: EDIPUCRS.
- Zilio, Leonardo. 2010. Terminologia textual e linguística de corpus: estudo em parceria. Em Cristina Becker Lopes Perna, Heloísa Orsi Koch Delgado & Maria José Bocorny Finatto (eds.), *Linguagens especializadas em corpora: modos de dizer e interfaces de pesquisa*, Porto Alegre: EDIPUCRS.