

**Gêneros textuais *datasheet* e artigo científico em aulas de ESP:
levantamentos léxico-estatísticos para fins educacionais**
**Datasheet and research article genres in ESP classes: lexical-statistic surveys for
educational purposes**

Sabrina Bonqueves Fadanelli*
Andrea Jessica Monzón**

RESUMO: A linha Sociorretórica de estudos de gêneros textuais (SWALES, 1990; BAZERMAN, 1997) tem apresentado resultados interessantes para o ensino de Inglês para Fins Específicos (ESP), muitos destes voltados para gêneros acadêmicos (BIBER et al., 2002). O presente artigo objetivou extrair dados relevantes para a preparação de material para o ensino de terminologia técnica em aulas de ESP no ambiente acadêmico. Assim sendo, artigos acadêmicos da área de Computação foram comparados a um gênero específico da área de Eletrotécnica: os *datasheets*. Utilizando ferramentas computacionais como o AntConc e o TermoStat e *corpora* destes dois gêneros, foram observadas as características apresentadas pelas ocorrências mais frequentes de colocações especializadas (HAUSMANN, 1990). Os resultados apontam similaridades e diferenças entre as colocações especializadas nos gêneros artigo acadêmico e *datasheet*, evidenciando a importância da aplicação da Linguística de *Corpus* no planejamento do professor de ESP.

PALAVRAS-CHAVE: Gêneros acadêmicos. Léxico especializado. Inglês para fins específicos. *Corpora*.

ABSTRACT: The social-rhetorical line of studies of text genres (SWALES, 1990; BAZERMAN, 1997) has presented interesting results in the teaching of English for Specific Purposes (ESP). A considerable amount of this research is related to academic genres (BIBER et al., 2002). This paper aims at extracting relevant data for the preparation of material for the teaching of technical terminology in ESP in the academic environment. Research articles in the area of Computer Science have been compared to a very specific genre in the area of Power Engineering: *datasheets*. Through the use of computing tools such as AntConc and TermoStat and *corpora* of both genres, features of the most frequent occurrences of specialized collocations (HAUSMANN, 1990) have been observed. The results point to similarities and differences between specialized collocations in the academic articles and the *datasheets*, evidencing the importance of applying *Corpus* Linguistics in the planning of ESP teaching.

KEYWORDS: Academic genres. Specialized lexicon. English for specific purposes. *Corpora*.

* Professora na área de Letras da Universidade de Caxias do Sul (UCS). Doutoranda do Programa de Pós-Graduação em Letras da Universidade Federal do Rio Grande do Sul (UFRGS). E-mail: sbfadane@ucs.br

** Professora de Português e Inglês no Instituto Federal do Rio Grande do Sul (IFRS) - Campus Feliz. Doutoranda do Programa de Pós-Graduação em Letras da Universidade Federal do Rio Grande do Sul (UFRGS). E-mail: andrea.monzon@feliz.ifrs.edu.br

1. Introdução

Nos últimos anos, a ciência tem assumido cada vez mais um caráter de força produtiva, haja vista a sua inserção no cotidiano das pessoas comuns. A atualização constante da tecnologia e o desaparecimento das fronteiras de comunicação por meio do advento da internet e das mídias em geral têm como uma das consequências a necessidade de se obter conhecimento sobre Língua Inglesa em praticamente todas as áreas do conhecimento. Isso porque essa língua é a língua franca (JENKINS, 2012) do conhecimento e das trocas de comunicação.

A Língua Inglesa é a principal linguagem da ciência mundialmente (CRYSTAL, 2010; GRADDOL, 2000), sendo pré-requisito na maioria dos cursos de pós-graduação e também em publicações de pesquisa em diversas revistas acadêmicas (PAIVA; PAGANO, 2001). Muitos cursos técnicos e de graduação também disponibilizam disciplinas de inglês em seus currículos. E, em paralelo, a linguagem técnica, mesmo em inglês, aparece cada vez mais em nosso dia a dia (FINATTO, 2011) no Brasil. Logo, tornou-se comum encontrar termos anglófonos em meio a textos técnico-científicos escritos em português, o que mostra a relevância de tal idioma nesse cenário.

O inglês voltado à compreensão da linguagem técnica de uma determinada área se insere no campo do ensino de ESP - *English for Specific Purposes* (Inglês para Propósitos ou Fins Específicos), uma abordagem em que o aluno recebe instrução para necessidades pessoais, acadêmicas e/ou profissionais específicas (HUTCHINSON; WATERS, 1987). Muitas vezes a disciplina nos institutos, faculdades e universidades também é chamada de Inglês Instrumental ou Inglês Técnico. Dois domínios técnicos que se enquadram neste cenário de necessidade de leitura em Língua Inglesa são a Ciência da Computação e a Eletrotécnica.

Na área de Ciência da Computação/Informática, para que alunos, docentes e profissionais se informem, capacitem e atualizem, os artigos científicos (SWALES, 1990) são os textos especializados mais respeitados. A maioria deles, entretanto, é publicada em inglês, o que demanda a leitura e compreensão de tal idioma, especialmente das terminologias de domínio.

A Eletrotécnica é a ramificação da Engenharia Elétrica responsável por instalações de redes elétricas residenciais e industriais. A principal necessidade de estudantes e profissionais para a realização de seu trabalho nesse domínio é a leitura de documentos técnicos chamados *datasheets* (DEWEY, 1998), os quais contêm especificações sobre o funcionamento de dispositivos elétricos. Esses documentos se encontram na internet, praticamente apenas em

Língua Inglesa e demandam um certo nível de proficiência na leitura (pelo menos um A2, de acordo com o *Common European Framework*¹).

Em se tratando de ensino-aprendizagem de ESP, segundo Bhatia, Anthony e Noguchi (2011), há 3 aspectos sobre os quais os professores e pesquisadores envolvidos com tal abordagem deveriam refletir: análise das necessidades, desenvolvimento de materiais (didáticos) e as práticas docentes. Algo que permeia esses aspectos é, tanto para discentes quanto docentes, conhecer diversos gêneros textuais com o intuito de participar de distintas comunidades discursivas (SWALES, 1990).

Desse modo, o presente trabalho tem o objetivo de identificar e analisar as características léxico-estatísticas dos gêneros textuais *datasheet* e artigo científico da Computação. Para tanto, aqui observamos as convencionalidades de tais textos sob duas perspectivas: 1) descrição linguística quantitativa e qualitativa através da compilação de *corpora* e do uso de duas ferramentas computacionais para as análises léxico-estatísticas, e 2) contribuições para as aulas de ESP em contexto de Ensino Técnico e Superior. Trata-se, aqui, de levantamentos e análises provenientes de duas teses de doutorado em fase de conclusão, as quais tratam da leitura em situação de ESP em cursos de Educação Profissional nas áreas de Computação e Eletrotécnica, sob as perspectivas teórico-metodológicas da Linguística Aplicada, Linguística de *Corpus* e Terminologia. Para o presente artigo, entretanto, não apresentamos e discutimos todos os resultados destas pesquisas, uma vez que aqui buscamos uma concentração nas análises lexicais dos *corpora* e suas possíveis contribuições linguístico-pedagógicas. Frisamos também que o gênero textual *datasheet* receberá uma descrição mais específica, não por ser mais relevante academicamente do que o gênero artigo acadêmico, mas justamente por ainda não ser tão estudado quanto o artigo acadêmico.

Assim, na seção 2, buscamos discutir aspectos teórico-metodológicos dos gêneros textuais e suas implicações para este trabalho. Já na seção 3, trazemos características e etapas metodológicas para o levantamento léxico-estatístico e, na seção 4, discutimos os resultados provenientes do mesmo, identificando as idiossincrasias da cada área do conhecimento aqui observada. Finalmente, na seção 5 apontamos o que aprendemos com este estudo, bem como indicamos potenciais caminhos a serem percorridos no futuro.

¹ Disponível em: http://www.coe.int/t/dg4/linguistic/Cadre1_en.asp

2. Os Gêneros Textuais e suas potencialidades teóricas, linguísticas e educacionais

Os gêneros textuais têm sido objeto de interesse por parte de estudiosos preocupados com uma prática que privilegie a interação verbal e as diversas situações de interlocução. Entretanto, nas palavras de Marcuschi (2008, p. 147), "o estudo dos gêneros textuais não é novo e, no Ocidente, já tem pelo menos vinte e cinco séculos, se considerarmos que sua observação sistemática se iniciou com Platão". Isso indica que, devido ao longo período de observações, o tema apresenta diversos enfoques teóricos ainda hoje.

A partir do século XIX, o estudo dos gêneros assumiu novas e importantes perspectivas com o surgimento da ciência da Linguagem, a Linguística. Por todo o século XX e início do XXI, houve o desenvolvimento de linhas de pensamento que enfatizam pontos de vista diferentes. As linhas de pensamento mais conhecidas, juntamente com seus principais expoentes, são (MEURER; BONINI; MOTTA-ROTH, 2005²):

- a) Sociodiscursiva (Bakhtin, Adam, Bronckart, Marcuschi, entre outros);
- b) Sociossemiótica (Hasan, Martin, Kress, Fowler entre outros);
- c) Sociorretórica (Swales, Bazerman, Bhatia, entre outros).

Mikhail Bakhtin formulou a teoria sobre gêneros que viria a influenciar o pensamento ocidental, com uma concepção de linguagem que incorporava às atividades de linguagem a visão de um sujeito que a utilizava como um instrumento de mudança social (BAKHTIN, 2003). Ele partiu do pressuposto de que a comunicação só se faz possível através de textos e/ou gêneros textuais. Destacando o lado social da noção de gênero, ele afirma que qualquer enunciado considerado isoladamente é individual, porém cada esfera de utilização da língua elabora seus tipos relativamente estáveis de enunciados, sendo por isso denominados gêneros do discurso.

Marcuschi (2003) transporta a linha de Bakhtin para a noção de gênero textual, afirmando que esses são fenômenos históricos profundamente vinculados à vida cultural e social. São, portanto, entidades sociodiscursivas e formas de ação social em qualquer situação comunicativa.

Nas abordagens de ordem Sociodiscursiva e Sociossemiótica, percebe-se que o gênero textual é abordado como um resultado de eventos e atividades humanas. Nesta pesquisa, preferimos adotar a abordagem Sociorretórica dos estudos de texto, seguindo principalmente as

² Os autores discorrem sobre as diferentes linhas teóricas referentes ao estudo de Gêneros ao longo de toda a obra.

linhas de Swales (1990) e Bazerman (2011). Isso se deve à característica impressa nesta abordagem de que o gênero textual é um agente que determina e modifica eventos e convenções sociais, ao invés de ser apenas visto como resultante dos mesmos. Bazerman (2011, p. 32) define gênero textual como:

fenômenos de reconhecimento psicossocial, que são parte de processos de atividades socialmente organizadas. Gêneros são tão-somente os tipos que as pessoas reconhecem como sendo usados por elas próprias e pelos outros. (...) Gêneros emergem nos processos sociais em que as pessoas tentam compreender umas às outras suficientemente bem para coordenar atividades e compartilhar significados com vistas a seus propósitos práticos.

Charles Bazerman (2005) entende o gênero como uma construção acionada pelo sujeito para reconhecimento e desenvolvimento de ações tipificadas em situações retóricas recorrentes, não como uma categoria linguística modelizada por certas características textuais. Esse direcionamento nos leva a perceber o gênero como um fato social, entendido como aquilo em que se acredita como sendo verdadeiro, afetando o modo como se definem as situações sociais. Cada texto bem-sucedido cria para seus leitores um fato social, consistindo em uma ação social significativa realizada pela linguagem.

Para Swales (1990), o gênero textual é, ainda, um conjunto de propósitos comunicativos formado pelo próprio discurso e seus participantes (comunidade discursiva). O propósito comunicativo para Swales é central em sua noção de gênero, visto que esse é a razão do uso da linguagem, fornecendo a justificativa por trás do gênero e contribuindo para a estruturação do discurso e das convenções de uso.

Há três conceitos intimamente relacionados: comunidade discursiva, propósito comunicativo e gênero. Esses são importantes na caracterização dos *datasheets* e dos artigos científicos da Computação, pois: (a) identificando quem os usa, (b) para que finalidade e (c) como os documentos se organizam, podemos mais adiante delinear contribuições mais eficientes para nossos contextos educacionais de ESP e às necessidades de nossos alunos.

Sobre os participantes da comunicação e seu posicionamento em relação às áreas de Elétrica e Computação, Swales afirma que a análise de um gênero textual deve iniciar identificando o seu propósito comunicativo dentro de uma comunidade discursiva. A comunidade discursiva que escreve os *datasheets* e Artigos Científicos da Computação (doravante ACCs) é, em sua maioria, altamente especializada. A princípio, o público-alvo de leitura dos *datasheets* e ACCs é de especialistas, contudo levamos aqui em consideração

aqueles que são especialistas em formação, ou seja, pesquisadores iniciantes e/ou estudantes durante a sua formação acadêmico-profissional. Esse cenário se adapta bem à noção expressa por Swales e Marcuschi de que é a comunidade discursiva que produz o gênero de acordo com suas necessidades. Assim sendo, no caso dos *datasheets* e ACCs, o reconhecimento do gênero dentro de sua comunidade discursiva pode estar ligado, dentre outros aspectos, ao nível de especialização que os autores de ambos os gêneros textuais empregam nas terminologias dos documentos. Hoffmann (1998; KILLIAN; FINATTO, 2015)³, em sua descrição estrutural interna de textos especializados, aponta como uma característica fundamental do texto especializado o nível de especialidade dos participantes da comunicação. Cabré (2008) e Freixa (2002) determinam que a ‘quantificação’ do nível de especialização se dá principalmente através da relação entre parâmetros comunicativos e textuais, principalmente a relação entre os interlocutores. Ciapuscio e Kuguel (2002) também evidenciam a importância do nível de especialização dos participantes da comunicação para uma tipologização do texto especializado. Tanto Ciapuscio e Kuguel como Hoffmann determinam que o texto especializado pode ter participantes mais ou menos especialistas na área.

Neste trabalho, observamos os gêneros textuais não somente pelas suas características, construção e níveis de especialidade, como também entendemos suas contribuições para o ensino-aprendizagem de línguas estrangeiras e ESP.

2.1 Gêneros textuais e ESP

Inglês é o idioma que tem o lugar de língua franca (JENKINS, 2012) na comunidade internacional acadêmica e científica, ocupando, ainda, a posição de *global language* (CRYSTAL, 2003) no cenário mundial dos negócios, tecnologia e do trabalho. É principalmente através de tal idioma que pesquisas são divulgadas e compartilhadas com estudantes, professores e profissionais de todo o mundo.

No caso específico do alunado dos cursos da área de Informática/Computação, onde se concentra a maioria das matrículas na educação profissional federal, para que sejam iniciados e inseridos em seu domínio de formação acadêmico-profissional, é preciso ler e compreender textos especializados em inglês. Sendo essa a língua predominantemente utilizada em

³ In: FINATTO, M. J.; ZILIO, L. **Textos e termos por Lothar Hoffmann** - um convite para o estudo das linguagens técnico-científicas. Porto Alegre: Palotti, 2015.

publicações científicas, livros, manuais e linguagens de computador, ela tem o status de ferramenta primordial para a assimilação e construção do “saber e [do] saber fazer específicos daquela ciência” (BARBOSA, 2009, p. 43).

Já no caso da área de Engenharia Elétrica, o artigo científico possui a sua devida importância, contudo no cotidiano acadêmico-profissional os alunos necessitam, em primeira instância, da leitura e compreensão dos *datasheets*. Nesse gênero textual são encontradas as terminologias em seu *habitat* natural (CABRE, 2009; 2002) para que os alunos, como futuros especialistas, possam vivenciar o uso e funcionamento de equipamentos elétricos recorrentes.

Para este contexto acadêmico, o da Educação Profissional, outra característica relevante dos gêneros textuais, apontada por Biber et al. (2004; 1995) é a sua composicionalidade. Existem padrões associativos que formam pacotes lexicais, os quais são “sequências recorrentes de palavras, que são importantes blocos constitutivos do discurso [acadêmico] nos registros falado e escrito”⁴ (BIBER, 2007, p. 1). O entendimento desses blocos de informações (SCARAMUCCI, 2007; 1998) só constrói seus sentidos através da leitura atenta, que não se atém a palavras isoladas, já que o significado de uma palavra “é sempre contextual” (FIRTH, 1957, p. 7).

No que tange à leitura, enquanto processamento, Koch (2005) defende a ideia de que os indivíduos desenvolvem uma competência metagenérica que lhes possibilita interagir com as línguas de forma conveniente. É “essa competência que possibilita a produção e compreensão de gêneros textuais” (KOCH; ELIAS, 2014, p. 102).

2.2 Colocações especializadas: blocos lexicais idiossincráticos

As colocações são fenômenos linguísticos arbitrários que são consagrados pelo uso ou pela prática, ou seja, a convencionalidade estabelece a sua combinabilidade. Esse conceito, por sua vez, remete “à faculdade que os elementos linguísticos têm de se combinar. Existem palavras que se associam de forma tão natural que a única explicação possível é de que essa associação tenha sido consagrada pelo uso, isto é, de que ela seja convencional” (TAGNIN, 2013, p. 25-26).

O estudioso que introduziu o termo *colocação* para as convencionalidades léxico-sintáticas foi J. R. Firth (1957). Propõe-se, aqui, estudar as colocações sob a ótica de que as

⁴ Tradução nossa.

palavras se constituem de acordo com a companhia que elas têm. As coocorrências podem ser arbitrárias, mas “se um de seus componentes for alterado, poderá haver ruído na comunicação” (SANTOS, 2010, p. 100).

Aqui optamos por empregar o termo colocações especializadas - CEs (ORENHA-OTTAIANO, 2012; L’HOMME; BERTRAND, 2000), devido à relevância terminológica e conceitual desses padrões lexicais dentro do domínio da Computação, como algo que remete às metalinguagens que devem ser inferidas pelos alunos em formação. O contato com essas terminologias ocorre através da leitura do inglês especializado.

As colocações, por sua vez, distinguem-se das combinatórias livres, nas quais as palavras se combinam por mera afinidade. Nas colocações, um usuário da língua pode entender as palavras que as compõem, mas não saberá usá-las e/ou reproduzi-las, não sendo essa tarefa algo automático. Dentro de uma colocação, o status de cada uma de suas partes não é igual. Hausmann (1990) denomina *base*⁵ (*keyword*) o componente que não modifica a identificação do que é caracterizado; já *collocatif* (colocado) é o componente que só recebe seu significado semântico através da construção da colocação em si. Por exemplo, na Computação o termo *display server* tem o primeiro vocábulo como um colocado (*display*), que com sua base (*server*) constrói um significado diverso do termo *server* por si só.

De acordo com L’Homme e Bertrand (2000, p. 498), “Colocações são convenções dentro de uma determinada comunidade linguística; as combinatórias lexicais especializadas são convenções dentro de um grupo de especialistas”.⁶ Essas autoras afirmam, ainda, que tanto dizer colocações especializadas quanto combinatórias léxicas especializadas atende a uma mesma comunidade de especialistas. Para não nativos, a importância de se conhecer e estudar colocações implica entender melhor algo que não pode ser inventado ou parafraseado (SANTOS, 2010). Essas convenções não podem ser vistas como algo com regularidade sintática ou semântica, uma vez que os aprendizes e/ou não nativos devem adquiri-las partindo do pressuposto de que elas são imprevisíveis (HAUSMMAN, 1990; L’HOMME; BERTRAND, 2000). Nisso reside a relevância de se produzir compilações de CEs de todas as áreas e subáreas do conhecimento. Pesquisadores e estudantes dos mais diversos níveis de proficiência em inglês

⁵ Texto original em francês.

⁶ Tradução nossa do trecho “*Collocations are conventional within a given linguistic community; specialized lexical combinations are conventional a group of specialists*”.

devem se comunicar de maneira produtiva em seus campos de estudo, divulgando e recebendo informações.

Para quem é aprendiz, portanto, a reprodução das colocações não ocorre automaticamente, ou seja, ele deve aprendê-las, “pois as línguas, dentro de sua totalidade de combinações lógicas possíveis, fazem escolhas idiossincráticas” (HAUSMMAN, 1990, p. 1010). Logo, a colocação é uma unidade, não da fala, mas da língua em si.

Seja em forma de repositórios, glossários, dicionários ou ferramentas pedagógicas, a arbitrariedade das CEs deve ser contemplada em produtos terminológicos que atendam a usuários reais com as dificuldades já discutidas.

Segundo Tagnin (2013), as colocações podem ser: nominais, adjetivas, verbais ou adverbiais. Sendo as CEs combinatórias que têm essencialmente cunho terminológico e que se configuram por uma combinabilidade convencionalizada por seus usuários especialistas, L’Homme (2000) menciona que sua base é majoritariamente nominal, havendo as seguintes possibilidades: verbo + substantivo, substantivo + adjetivo e substantivo + substantivo. Orenha-Ottaiano (2012), entretanto, discorda de que haja somente essas três possibilidades para CEs, uma vez que a pesquisadora verificou, em seus estudos colocacionais acerca de jornalismo de negócios e contratos sociais em textos anglófonos, que a taxonomia de Hausmann (Quadro 1) pode ser empregada tanto para analisar colocações da língua comum quanto da de especialidade. Assim sendo, essa foi a classificação aqui aplicada.

Quadro 1 — Taxonomia de Hausmann para as colocações.

colocações nominais	substantivo + substantivo substantivo + preposição + substantivo
colocações verbais	verbo + substantivo substantivo + verbo verbo + preposição + substantivo verbo + partícula adverbial + substantivo verbo + adjetivo
colocações adjetivas	adjetivo + substantivo
colocações adverbiais	advérbio + adjetivo verbo + advérbio advérbio + verbo

Fonte: Hausmann, 1985 (apud ORENHA-OTTAIANO, 2012, p. 153).

Percebemos a relevância das CEs, enquanto blocos de informações convencionalizados, ao longo dos dois gêneros textuais aqui analisados - *datasheet* e artigo científico da Computação. Isso demonstra a potencialidade de se utilizar estudos linguístico-descritivos, quantitativos e qualitativos, no âmbito do léxico técnico-profissional como ferramentas pedagógicas e de aporte teórico-metodológico para aulas de ESP, as quais trabalham fundamentalmente com gêneros especializados.

2.3 Gênero textual *datasheet* e suas peculiaridades

Apresentamos a seguir algumas características macro e microestruturais de *datasheets* de Eletrotécnica. O Quadro 2, de nossa autoria juntamente com um especialista engenheiro elétrico, mostra os aspectos macroestruturais que os *datasheets* em geral apresentam. Um exemplo de *datasheet* pode ser visualizado na Figura 1.

Quadro 2 — Características em comum dos *datasheets*.

característica	lâmpada led	inversor fotovoltaico	capacitor
nome da empresa	X	X	X
endereço, site e contato		X	X
modelo/código do produto	X	X	X
aplicações	X	X	X
desenhos, dimensões, diagramas	X	X	X
gráficos	X	X	
condições de funcionamento (tensão, consumo, temperaturas, etc.)	X	X	X
características de padrão, elétricas e ópticas	X	X	X
símbolos, referências numéricas, abreviações	X	X	X

Fonte: elaborado pelas autoras.

A microestrutura dos *datasheets* pode mesclar características descritivas/argumentativas e expositivas: os dados técnicos e informações de atributos do produto são colocados objetivamente em tabelas ou introduzidas por marcadores. Aparecem fotos ou desenhos do produto e as informações de contato da empresa fabricante, além de gráficos. O documento pode apresentar um *layout* colorido ou preto-e-branco. *Datasheets* podem ser usados como ferramentas de venda do produto, como também podem acompanhar o mesmo no processo de pós-venda, a fim de fornecer dados para sua instalação e manutenção.

Figura 1 — Exemplo de *datasheet*

World Components Network Service Ltd.	Model No.	WCN-411HY1-110N
	Doc. No.	LED-411HY1-110N

Applications:

- Advertising Signs
- Indicators
- LCD Back Light
- Illuminations

Absolute Maximum Ratings at Ta = 25°C

Items	Symbol	Absolute maximum Rating	Unit
Forward Current	I_F	50	mA
Peak Forward Current*	I_{FP}	200	mA
Reverse Voltage	V_R	5	V
Power Dissipation	P_D	130	mW
Operation Temperature	T_{opr}	-40 ~ + 95	°C
Storage Temperature	T_{sta}	-40 ~ + 100	°C
Lead Soldering Temperature	T_{sol}	Max.260°C for 3 sec Max. (3mm from the base of the epoxy bulb)	

pulse width $\leq 0.1\text{msec}$ duty $\leq 1/10$

Typical Electrical & Optical Characteristics (Ta = 25°C)

Items	Symbol	Condition	Min.	Typ.	Max.	Unit
Forward Voltage	V_F	$I_F = 20\text{mA}$	---	2.1	2.6	V
Reverse Current	I_R	$V_R = 5\text{V}$	---	---	100	μA
Luminous Intensity	I_V	$I_F = 20\text{mA}$	---	700	---	mcd
Peak Emission Wavelength	• • •	$I_F = 20\text{mA}$	---	---	---	nm
Dominant Wavelength	• • •	$I_F = 20\text{mA}$	585	591	595	nm
Spectral Line Half-Width	• • • •	$I_F = 20\text{mA}$	---	26	---	nm
50% Power Angle	$2\theta_{\frac{1}{2}}$	$I_F = 20\text{mA}$	---	100/40	---	deg

Dimension Drawing

NOTES:
1. ALL DIMENSIONS ARE IN mm TOLERANCE IS +/- 0.25, UNLESS OTHERWISE NOTED.
2. BURR AROUND BOTTOM OF EPOXY MAY BE 0.5mm MAX.

Important Notes:

- 1) All ranks will be included per delivery, rank ratio will be determined by WCN.
- 2) Tolerance of measurement of luminous intensity is $\pm 15\%$.
- 3) Tolerance of measurement of Vf is $\pm 0.05\text{V}$.
- 4) Packaging methods are available for selection, please refer to PACKAGING STANDARD.
- 5) Please refer to LED LAMP RELIABILITY TEST STANDARD for reliability test conditions.

3. Metodologia

A Linguística de *Corpus* (LC) é uma abordagem que possui como característica principal a análise empírica, a partir de padrões existentes em textos naturais, fazendo uso de ferramentas computadorizadas e dependendo de enfoques qualitativos e quantitativos (BERBER SARDINHA, 2004). Diferentemente das abordagens mais tradicionais, que determinam que a linguagem deve ser analisada de acordo com princípios determinados anteriormente, a LC primeiro observa o comportamento da linguagem em seu *habitat* natural (o texto) e depois teoriza sobre ela. Para esta abordagem da linguagem, existe uma correlação entre as características linguísticas e contextuais, podendo-se observar mais facilmente padrões na língua sob análise⁷.

3.1 Compilação de *corpora*

O *corpus* de estudo para observação dos *datasheets* foi coletado a partir de documentos de componentes elétricos e áreas da Eletrotécnica sugeridas por dois engenheiros elétricos, de acordo com as necessidades mais proeminentes de leitura que os alunos desta área apresentam. Os *datasheets* podem ser facilmente encontrados na internet e podem ser baixados em formato eletrônico. Assim, separamos os textos em 11 componentes/áreas pertinentes à Eletrotécnica divididas em 11 arquivos txt, os quais limpamos conforme os passos propostos por Aluísio e Almeida (2006) removendo letras maiúsculas, números, pontuação e símbolos. Não removemos, contudo, as tabelas e figuras, uma vez que essas contêm informações extremamente relevantes e idiossincráticas dos *datasheets*. O total de *tokens* foi 29.718. Os componentes e áreas dos *datasheets* são os seguintes: capacitores, conversores, diodos, disjuntores, indutores, inversores fotovoltaicos, osciladores de cristal, resistores, transistores, transformadores e segurança.

Para a observação dos artigos científicos da Computação, foi compilado um *corpus* de cerca de 400.000 *tokens*, contendo 54 artigos científicos da Computação provenientes de seis periódicos da Association of Computer Machinery (ACM), datados de 2010 a 2014. Os títulos

⁷ Reiteramos que para que as escolhas metodológicas fossem constituídas, foram realizados estudos-piloto (MONZÓN; FADANELLI, 2016) com *corpora* e com experimentos de leitura com os públicos-alvo discentes destas teses de doutorado.

dos periódicos foram selecionados por dois professores especialistas da área de Informática, os quais lecionam em cursos técnicos e tecnológicos (Ensino Superior) de tal domínio.

3.2 Ferramentas de PLN para estudos léxico-estatísticos

As ferramentas extratoras utilizadas no estudo foram de natureza estatística e linguística, com o intuito de promover mais variedade nos resultados (VIEIRA; LOPES, 2010); dificilmente um extrator de candidatos a termos obtém resultados 100% confiáveis. Como exemplo podemos citar Teixeira (2010), que em seu estudo comparou quatro ferramentas de extração de termos diferentes e constatou que nenhum obteve mais do que 20% de índice de acerto. As seguintes ferramentas computacionais foram escolhidas por serem de fácil uso e obtidas de forma gratuita na Internet: o *AntConc 3.2.4*, recurso que se utiliza mais de estatística de frequência; e o *TermoStat*, recurso que usa estatística linguística.

O AntConc é um *kit* de ferramentas de análise de *corpus* desenhado por Laurence Anthony (ANTHONY, 2005; 2004). É equipado com um concordanciador⁸, geradores de listas de frequência de palavras e a ferramenta utilizada para a extração, a função *Keyword* (oferece na tela a lista da candidatas a palavras-chave de um *corpus*, comparando o *corpus* de estudo com um outro *corpus* no mínimo cinco vezes maior - chamado de *corpus* de referência); o *corpus* de referência foi composto de textos extraídos do *corpus* COCA (*Contemporary Corpus of American English*) e do BNC (*British National Corpus*), pesquisando a ocorrência das 30 palavras mais frequentes na wordlist do *Corpus* COCA. O *corpus* de referência possui aproximadamente 211 mil *tokens*, mais do que cinco vezes o tamanho do *corpus* de estudo (SARDINHA, 2000). Os textos têm origens de diversos domínios, acadêmico, noticiários, literatura, internet, e passaram pela mesma limpeza de pontuação, etc. que o *corpus* de pesquisa.

O TermoStat (DROUIN, 2003) é um extrator de termos acessado gratuitamente na *web*, que primeiramente se baseia em outro programa, o *TreeTagger* (SCHMID, 1994) para *tokenizar*, lematizar e etiquetar gramaticalmente o *corpus*. Depois deste processo, o programa junta candidatos a termos (unigramas e bigramas) em uma lista de acordo com padrões gramaticais. O TermoStat fornece outras informações além da frequência, como por exemplo a

⁸ Um concordanciador, de acordo com Sardinha (2009, p. 9), é “uma ferramenta que: realiza concordâncias, ou listagens de uma palavra específica (ou nóculo) juntamente com parte do texto onde ocorreu. Oferece também listas de colocados, isto é, palavras que ocorreram perto do nóculo”.

classificação morfofossintática dos termos, relações de combinabilidade e grafos de relações semânticas.

Essas duas ferramentas se complementam quando se trata de identificar e analisar padrões lexicais nos *corpora*, especialmente sob a ótica terminológica, a qual se sobressai nestes textos produzidos por especialistas. Salientamos, entretanto, que a seleção automática é complementada pela análise manual.

4. Resultados e discussão

Apesar das particularidades e convencionais da Eletrotécnica e da Computação, no que diz respeito à composicionalidade lexical observada em gêneros textuais relevantes para as comunidades de cada um desses domínios, existem aspectos semelhantes. Desse modo, nesta seção analisaremos os padrões léxico-estatísticos neste contexto, atentando para suas potenciais contribuições para o ensino-aprendizagem de ESP na Educação Profissional.

4.1 Padrões léxico-estatísticos em artigos científicos da Computação

Desse modo, ao se levantarem semiautomaticamente as palavras gramaticais mais frequentes, utilizou-se o AntConc em parceria com o TermoStat, para então se fazer uma lista de termos que são candidatos a serem bases (HAUSMANN, 1990) para CEs; obtiveram-se resultados um pouco diferentes através dessas ferramentas (Tabelas 1 e 2). É importante salientar, ainda, que como o TermoStat calcula o *score* de especificidade de cada termo (*score spécificité*), sua lista acaba trazendo informações complementares em relação ao AntConc. Mesmo assim, a análise é semiautomática⁹, pois, neste estudo, sempre houve o julgamento humano ao percorrer as listas linha por linha e “limpá-las” de forma a garantir a confiabilidade dos termos. Isso porque o levantamento e a análise exclusivamente automáticos trazem alguns equívocos bastante visíveis como não sendo de linguagem natural.

⁹ “A Linguística de *Corpus* tem se ocupado, dentre outras coisas, de criar procedimentos automatizados para o levantamento da frequência lexical, mas atualmente ainda é necessária a intervenção humana para apurar os resultados obtidos pelos programas de consulta aos *corpora*” (DURAN, XATARA, 2006, p. 44).

Tabela 1 — Lista dos 10 candidatos a termos (unigramas) mais recorrentes na *wordlist* do AntConc.

termo	frequência
time	965
set	754
students	749
algorithm	701
problem	692
user	584
use	578
data	539
number	539
learning	526

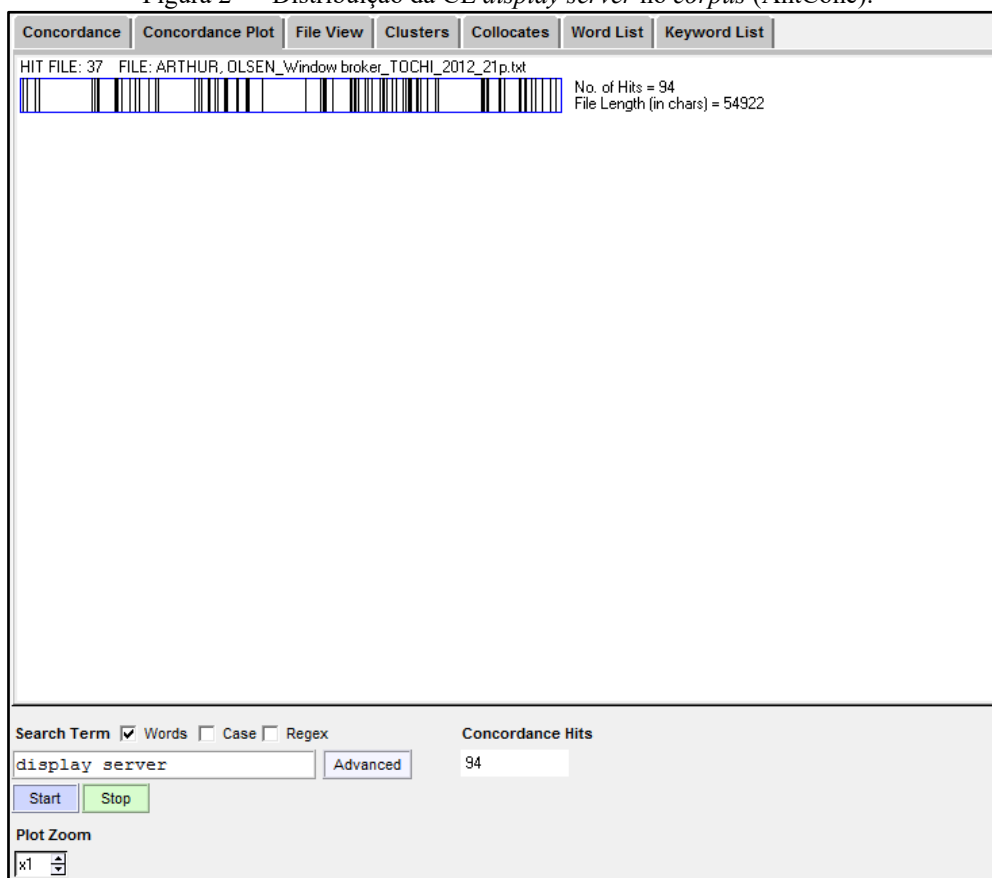
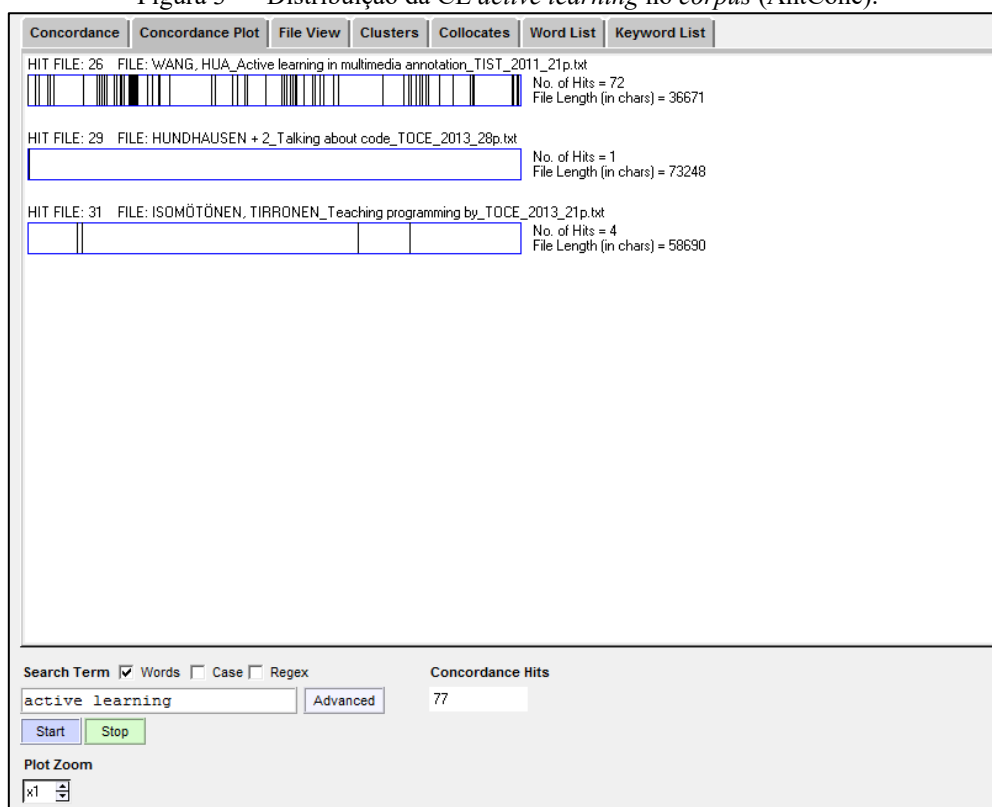
Tabela 2 — Lista dos 10 termos (unigramas) mais recorrentes na lista do TermoStat.

termo	frequência	score spécificité
algorithm	886	123.21
use	962	114.28
graph	561	94.04
tag	499	86.16
participant	539	82.79
vertices	284	70.23
set	685	69.13
edge	507	68.11
node	298	67.74
model	628	66.81

Outro julgamento e análise necessários à seleção de CEs foi verificar se cada uma das candidatas (Quadro 3) ocorria em pelo menos três artigos científicos e gerar uma lista mais apurada (Quadro 4). No Quadro 3 estão apresentadas as 17 CEs mais frequentes geradas pelo TermoStat. A primeira delas é *display server*, a qual foi possível verificar, através do AntConc, que está presente em somente um dos textos do *corpus* (Figura 2). Através da funcionalidade denominada *concordance plot* visualiza-se a distribuição de uma CE, ou qualquer outro tipo de *cluster*, ao longo dos textos em que a ocorrência é observada (Figura 3). Dessa forma, a primeira CE do Quadro 3, ao ser analisada, é encontrada em apenas um artigo do *corpus*. Consequentemente, essa CE foi extraída da lista final. O mesmo ocorreu com *information goal*, sendo que *active learning* foi a primeira CE da lista que ocorre em vários textos. Nesse caso, o termo ocorre em três textos, sendo que é bem mais frequente em um deles (Figura 3). Finalmente, chegou-se a uma listagem de 7 CEs (Quadro 4), eliminando-se 10 delas da lista inicial.

Quadro 3 — Lista de 17 CEs geradas pelo TermoStat.

CEs
display server
information goal
fairness property
surrogate competition
display space
active learning
window broker
social network
user interface
nominal group
utility function
datum structure
relational structure
user interest
tag choice
broker software
high probability
social influence

Figura 2 — Distribuição da CE *display server* no corpus (AntConc).Figura 3 — Distribuição da CE *active learning* no corpus (AntConc).

Quadro 4 — Lista de 7 CEs depois da seleção manual.

CEs selecionadas
active learning (adj. + substantivo)
social network (adj. + substantivo)
user interface (substantivo + substantivo)
high propability (adj. + substantivo)
planar graph (adj. + substantivo)
polynomial time (adj. + substantivo)
data structure (substantivo + substantivo)

Levantou-se de forma automática que 31% das CEs são colocações adjetivas (substantivo + adjetivo)¹⁰ e 27% são nominais (substantivo + substantivo). Ademais, atestou-se que 24% dos termos são unigramas compostos por um substantivo. Logo, totaliza-se nessas três composições 82% dos termos listados pelo TermoStat (Tabela 3).

Tabela 3 — Composição dos termos listados pelo TermoStat.

composição dos termos	frequência	porcentagem
adjetivo + substantivo	1.843	31%
substantivo + substantivo	1.586	27%
substantivo	1.417	24%
adjetivo + substantivo + substantivo	180	3%
substantivo + substantivo + substantivo	129	2%
adjetivo + adjetivo + substantivo	76	1%
	5.231	88%

Para melhor compreender as CEs de forma mais aprofundada, bem como o universo das terminologias da Computação, foi importante observar atentamente o comportamento dos termos que as constituíam como base (*base, node*). Para demonstrar como esses levantamentos são obtidos ao se submeter este *corpus* ao TermoStat, organizou-se a Tabela 4.

¹⁰ Reitera-se que Hausmann (1990), em sua tipologia de colocações, determina que uma colocação adjetiva é formada por substantivo + adjetivo, porém na Língua Inglesa essas posições se invertem.

Tabela 4 — Características dos 10 termos mais frequentes como base de CEs (TermoStat).

Lema	frequência	score <i>spécificité</i>	variantes ortográficas	classe
algorithm	886	123.21	algorithm algorithms	substantivo
use	962	114.28	user users	substantivo
graph	561	94.04	graph graphs	substantivo
tag	499	86.16	tag tags	substantivo
participant	539	82.79	participant participants	substantivo
vertices	284	70.23	vertices	substantivo
set	685	69.13	set sets	substantivo
edge	507	68.11	edge edges	substantivo
node	298	67.74	node nodes	substantivo
model	628	66.81	model models	substantivo

Além disso, analisar as possibilidades combinatórias de um termo para gerar CEs é algo que agregou a este estudo, posto que as terminologias estabelecem as relações conceituais de uma determinada ciência (BARBOSA, 2009). Desse modo, quando se examinou o comportamento do termo (unigrama) mais frequente - *algorithm* -, notou-se uma gama bem interessante de relações de combinabilidade levantadas e quantificadas estatisticamente pelo TermoStat (Figura 4). Salientamos, portanto, que somente as análises automáticas não nos trazem todas as ocorrências relevantes nos *corpora* pensando-se especialmente nas aulas de ESP. Desse modo, muito embora o termo *algorithm* componha uma CE em 30^a posição na listagem automática e 10^a posição na lista limpa manualmente, acreditamos que se trata de um conteúdo léxico-conceitual relevante para o contexto deste estudo. Observou-se, então, que o adjetivo *deterministic*, por exemplo, é o que apresenta o maior índice de especificidade ao coocorrer com *algorithm*. Investigando-se a aproximação desses vocábulos ao constituir o termo *deterministic algorithm*, observa-se que eles mantêm uma relação semântica relevante, uma vez que se trata, de acordo com o especialista consultado, de um tipo específico de algoritmo, o que pode ser observado na Figura 4.

Figura 4 — Combinatórias envolvendo o termo *algorithm*.

Termes en relation	deterministic (191.84) approximation (79.86) polynomial-time (63.68) linear-time (63.68) efficient (57.84) dynamic (36.52) greedy (34.7) randomize (28.91) multiplier (24.17) latter (23.13) state-of-the-art (23.13) whole (23.13) layout (21.41) k-stroll (9.21) first (7.66)
Inclus dans	approximation algorithm broker algorithm deterministic algorithm dynamic algorithm efficient algorithm first algorithm greedy algorithm k-stroll algorithm latter algorithm layout algorithm linear-time algorithm multiplier algorithm polynomial-time algorithm randomized algorithm state-of-the-art algorithm time algorithm

4.2 Padrões léxico-estatísticos em *datasheets*

Para obter as classes morfossintáticas que estão mais presentes nos *datasheets* usou-se o etiquetador *Claws Free Tagger* da Universidade de Lancaster, Reino Unido. O etiquetador aceita arquivos com a extensão txt e classifica cada palavra com códigos como NN1 para substantivos, por exemplo. Foram inseridos no etiquetador os termos resultantes dos extratores AntConc. Os termos apontados pelo TermoStat não precisaram ser inseridos, pois o software já classifica morfologicamente os resultados. Como sabemos que este tipo de ferramenta não é 100% confiável por si só, um exame da etiquetagem foi feito e algumas classificações foram corrigidas manualmente. Após uma contagem das ocorrências de cada classificação morfossintática, as categorias que mais se destacaram, assim como nos ACCs, foram os substantivos, seguidos pelos adjetivos.

Segundo estudos previamente realizados em relação às características morfossintáticas de conceitos especializados, sabe-se que a grande maioria dos termos pertence à categoria dos substantivos (KRIEGER; FINATTO, 2004, p. 81). A alta quantidade de termos classificados como substantivos combina com o tipo de texto expositivo dos *datasheets*, cujo objetivo é apresentar peças e equipamentos elétricos e seus componentes.

A segunda categoria mais frequente foi a dos adjetivos, e aqui podemos notar a concordância com a característica de muitos *datasheets* que possuem textos descritivos e

argumentativos em sua estrutura, onde os adjetivos desempenham um papel importante na linha textual.

O mesmo processo de julgamento para a escolha de CEs feito com os artigos científicos da área da computação foi aplicado para resultar tanto nos termos mais frequentes (Tabelas 5 e 6), como na composição morfossintática dos CEs.

Tabela 5 — Lista dos 10 termos mais recorrentes no AntConc.

Termo	Frequência
output	245
power	227
voltage	220
current	157
temperature	151
circuit	147
data	124
capacitors	123
input	115
high	102

Tabela 6 — Lista dos 10 termos mais recorrentes no TermoStat.

Candidato	Frequência	Variantes
output	260	output__outputs
voltage	249	voltage__voltages
capacitor	230	capacitor__capacitors
power	214	power
circuit	163	circuit__circuits
temperature	154	temperature__temperatures
input	127	input__inputs
converter	110	converter__converters
operation	110	operation__operations
product	105	product__products

Na extração da composição morfossintática dos CEs através do AntConc, os resultados excluíram os CEs que constituíssem o nome fantasia do aparelho ou dispositivo elétrico descrito pelos *datasheets*. A Tabela 11 mostra a listagem dos 17 CEs mais frequentes e que ocorrem em ao menos três textos.

Tabela 7 — Listagem dos 17 CEs mais frequentes no AntConc.

CE	Frequência	Em quantos datasheets aparece
Output voltage (noun + noun)	54	7
Single output (adj + noun)	26	4
Input voltage (noun + noun)	25	7
Short circuit (adj + noun)	24	7
Power factor (noun + noun)	23	4
Output power (noun + noun)	22	7
Voltage range (noun + noun)	22	4
Power supply (noun + noun)	21	6
Output current (noun + noun)	18	5
Supply voltage (noun + noun)	17	7
Phase noise (noun + noun)	16	4
Rated voltage (adj + noun)	15	6
Ac dc (noun + noun)	12	3
Frequency stability (noun + noun)	12	3
Silicon power (noun + noun)	10	3
Circuit current (noun + noun)	9	6
Power dissipation (noun + noun)	7	4

No TermoStat, as CEs geradas são parecidas (seguindo os mesmos critérios de exclusão de nomes de aparelhos ou dispositivos), com pequenas diferenças provavelmente devido à natureza sintático-estatística do TermoStat (Tabela 8).

Tabela 8 — Listagem dos 17 CEs mais frequentes no TermoStat.

CE	Frequência
output voltage (noun + noun)	59
single output (adj + noun)	26
short circuit (adj + noun)	24
input voltage (noun + noun)	21
power factor (noun + noun)	19
output power (noun + noun)	17

phase noise (noun + noun)	16
power supply (noun + noun)	13
voltage range (noun + noun)	14
supply voltage (noun + noun)	14
rated voltage (adj + noun)	12
frequency stability (noun + noun)	12
output current (noun + noun)	11
ac dc (noun + noun)	10
silicon power (noun + noun)	8
circuit current (noun + noun)	8
power dissipation (noun + noun)	5

As CEs sintaticamente classificadas pelos extratores apresentam nos *datasheets* uma maioria de ocorrências do padrão substantivo + substantivo, embora haja ocorrências de adjetivo + substantivo. As implicações destes resultados e dos resultados mostrados na seção sobre a análise dos artigos acadêmicos serão comentadas a seguir.

4.3 Lições aprendidas com a comparação dos dois gêneros textuais

Se olharmos para a composição morfossintática dos CEs mais frequentes nos *datasheets* apontados por ambos extratores, veremos que quase todos pertencem à combinação substantivo + substantivo. Essa característica se relaciona com a microestrutura dos *datasheets*, que apresenta uma descrição e exposição de um produto, a fim de informar sobre seus componentes; porém, existe uma representação muito relevante do padrão adjetivo + substantivo na composição dos CEs, evidenciando uma outra possibilidade de uso dos *datasheets*: usados para fazer propaganda e vender o dispositivo.

Outro ponto importante a ser considerado diz respeito ao significado e relevância das CEs mais frequentes. Assim como no caso dos artigos científicos de Computação (rever exemplo sobre o termo *algorithm* em 4.1), as colocações não parecem ser aleatórias, e sim determinam conceitos bem específicos que fazem parte dos componentes elétricos. Os termos *output voltage/input voltage*, por exemplo, são dois tipos diferentes de tensão elétrica (saída e entrada). Outra característica interessante é que muitos dos mesmos termos individuais se combinam entre si para formar conceitos diferentes (*output voltage; output power; output*

current). E muitos termos individuais se posicionam de forma diferente na CE, e, embora sejam substantivos, acabam por assumir uma função adjetiva no significado da CE. Por exemplo: *supply voltage*, onde *voltage* é um substantivo; e *voltage range*, em que *voltage* assume uma função mais adjetivada.

A implicação mais evidente destes resultados é de que, embora sintaticamente a constituição dos CEs presentes em artigos acadêmicos e *datasheets* pareça se apresentar diferente, quando funcionalmente inseridas no gênero textual, os padrões substantivo + substantivo e adjetivo + adjetivo parecem se equilibrar.

Outra implicação muito importante, mais especificamente sobre os *datasheets*, é de que na verdade o vocabulário a ser trabalhado pelo professor não precisa ser muito vasto, já que as CEs mais frequentes se combinam entre si para formar conceitos diferentes. O significado isolado dos termos principais deve ser trabalhado como uma introdução para depois se enfatizar as combinatórias e os conceitos que as CEs expressam.

Já no caso dos ACCs, as CEs apresentam uma gama de vocabulário mais variado, significando que neste domínio científico o trabalho com os termos individuais e com as combinatórias deve receber a mesma atenção.

5. Conclusão

Em vista dos dados apresentados nas seções anteriores, foi possível verificar que a constituição de CEs entre os artigos acadêmicos de Computação e os *datasheets* de Eletrotécnica apresentam-se diferentes à primeira vista: nos artigos acadêmicos da área da Computação as combinações adjetivo + substantivo aparecem com maior frequência na escolha dos CEs que apareciam em três ou mais artigos acadêmicos. Já nos *datasheets*, a combinação substantivo + substantivo é mais recorrente nos CEs mais frequentes, sendo que a ocorrência do padrão adjetivo + substantivo também aparece, porém em menor escala.

No entanto, quando olhamos para as CEs e seu funcionamento dentro do gênero textual a que pertencem, percebemos que muitas das CEs que nos *datasheets* haviam sido classificadas como substantivos, se posicionam funcionalmente como adjetivos.

Todas estas descobertas são importantes para o professor de ESP que deseja produzir material e preparar aulas que realmente atendam às necessidades reais de seus alunos. Certamente, um destaque especial deve ser dado ao tratamento dos CEs nas diferentes áreas,

sempre chamando a atenção do aprendiz para a conexão que se estabelece entre os padrões linguísticos e o gênero textual com a qual se está lidando.

Através do levantamento automático, utilizando-se as ferramentas AntConc (ANTHONY, 2005; 2004) e TermoStat (DROUIN, 2003), verificou-se que as CEs desempenham um papel importante no cenário terminológico dos textos de ambos *corpora* pesquisados. As CEs mais frequentes de ambos gêneros textuais mencionados neste trabalho representam um exemplo muito importante de elementos a serem incluídos em atividades de aprendizagem e trabalho com uma língua especializada: os exemplos citados, com suas respectivas combinatórias, foram apontados por especialistas das áreas de Eletrotécnica e Computação como conceitos essenciais para o estudo das áreas em questão.

Por fim, percebemos que o escopo deste estudo ainda apresenta inúmeras possibilidades de expansão, através da análise de diferentes gêneros textuais de diversos domínios científicos, com o objetivo de profissionalizar a montagem de ferramentas, aulas e material didático para o ensino da Língua Inglesa com propósitos específicos (ESP).

Referências bibliográficas

ALUÍSIO, S.; ALMEIDA, G. O que é e como se constrói um *corpus*? Lições aprendidas na compilação de vários *corpora* para pesquisa linguística. **Calidoscópico**, São Leopoldo/RS, v. 4, n. 3, p. 155-177, 2006.

ANTHONY, L. AntConc: a learner *corpus* and a classroom friendly multi-platform *corpus* analysis toolkit. **Proceedings of Interactive Workshop on Language E-Learning**, Tokyo/Japan, p. 7-14, 2004.

ANTHONY, L. AntConc: design and development of a freeware *corpus* analysis toolkit for the technical writing classroom. **Proceedings of International Professional Communication Conference IPCC**, pp. 729-737, 2005. <https://doi.org/10.1109/ipcc.2005.1494244>

BAKHTIN, M. **Estética da criação verbal**. São Paulo: Martins Fontes, 2003.

BARBOSA, M. A. Terminologia aplicada: percursos interdisciplinares. **Polifonia** (UFMT), v. 4, p. 29-44, 2009.

BAZERMAN, C. **Gêneros textuais, tipificação e interação**. São Paulo: Cortez, 2005.

BIBER, D. The comprehensive analysis of register variation. In: **Dimensions of register variation: a cross-linguistic comparison**. Cambridge University Press, 1995. <https://doi.org/10.1017/CBO9780511519871.002>

BIBER, D.; CONRAD, S.; REPPEN, R.; BYRD, P.; HELT, M. Speaking and writing in the University: A multidimensional comparison. **TESOL Quarterly**, v.36, p. 9-48, 2002. <https://doi.org/10.2307/3588359>

BIBER, D.; CONRAD, S.; CORTES, V. If you look at...: lexical bundles in university teaching and textbooks. **Applied Linguistics**, v. 25, n. 3, p. 371-405, 2004. <https://doi.org/10.1093/applin/25.3.371>

CABRÉ, M. T. La Teoría Comunicativa de la Terminología: una aproximación lingüística a los términos. **Revue Française de Linguistique Appliquée**, v. 14, n. 2, p. 9-15, 2009.

CABRÉ, M. T. Textos especializados y unidades de conocimiento: metodología y tipologización. In: GARCÍA PALACIOS, J. and FUENTES, M. T. (Eds.). **Texto, terminología y traducción**. Salamanca: Ediciones Almar, 2002.

CIASPUCIO, G.; KUGUEL, I. Hacia una tipología del discurso especializado: aspectos teóricos y aplicados. In: PALACIO, J.G. e FUENTES, M.T. (eds.). **Entre la terminología, el texto y la traducción**. Salamanca, Almar, p. 37-73, 2002.

DEWEY, F. R A Complete Guide to Datasheets. **Sensors Magazine**, 1998. Disponível em: <http://www.allegromicro.com/~media/Files/Technical-Documents/pub26000-Complete-Guide-To-Datasheets.ashx>. Acesso: 01 dez. 2014.

DROUIN, P. Term extraction using non-technical *corpora* as a point of leverage. **Terminology**, v. 9, n. 1, p. 99-117, 2003. <https://doi.org/10.1075/term.9.1.06dro>

DURAN, M.; XATARA, C. A metalexicografia pedagógica. **Cadernos de Tradução (UFSC)**, v.2, n. 18, p. 41-66, 2006.

FADANELLI, S. Arquitetura de um glossário terminológico Inglês-Português na área de Eletrotécnica: rumo a uma Terminografia Didático-Pedagógica. **Anais do III Congresso Internacional Linguagem e Interação**, São Leopoldo/RS, 2015, p.165-166.

FINATTO, M. J.; ZILIO, L. **Textos e termos por Lothar Hoffmann** - um convite para o estudo das linguagens técnico-científicas. Porto Alegre: Palotti, 2015.

FIRTH, J. R. **Papers in Linguistics 1934-1951**. London: Oxford University Press, 1957.

HAUSMANN, F. J. Le dictionnaire de collocations. In: HAUSSMAN, F. J. [et al.]. **An international encyclopedia of lexicography**. V. 1. Berlin, New York: Walter de Gruyter, 1990, p. 1010-1019.

HOFFNAGEL, J. C.; DIONÍSIO, Â. P. (Orgs.). **Gêneros textuais, tipificação e interação**. São Paulo: Cortez, 2011.

HUTCHINSON, T.; WATERS, A. **English for Specific Purposes**. Cambridge University Press, 1987. <https://doi.org/10.1017/CBO9780511733031>

JENKINS, J. English as a *lingua franca*: from the classroom to the classroom. **ELT Journal**, Special Issue, p. 486-494, 2012. <https://doi.org/10.1093/elt/ccs040>

KOCH, I.; ELIAS, V. M. **Ler e compreender**: os sentidos do texto. São Paulo: Contexto, 2014.

KOCH, I. **O texto e a construção dos sentidos**. São Paulo: Contexto, 2005.

L'HOMME, M. C. Understanding Specialized Lexical Combinations. **Terminology**, v. 6, n. 1, p. 89-110, 2000. <https://doi.org/10.1075/term.6.1.06hom>

L'HOMME, M. C.; BERTRAND, C. Specialized Lexical Combinations: Should they be described as Collocations or in Terms of Selectional Restrictions? **Proceedings of Euralex 2000** - European Association of Lexicography, Stuttgart-Germany August 8th-12th.

MARCUSCHI, L. A. Gêneros textuais: definição e funcionalidade. In: DIONÍSIO, A. P.; MACHADO, A.R.; BEZERRA, M.A. (orgs.). **Gêneros Textuais e Ensino**. Rio de Janeiro: Lucena, 2003.

MARCUSCHI, L. A. **Produção textual, análise de gêneros e compreensão**. São Paulo: Parábola Editorial, 2008.

MEURER, J. L.; BONINI, A.; MOTTA-ROTH, D. **Gêneros. Teorias, métodos, debates**. São Paulo: Parábola Editorial, 2005.

MONZÓN, A.; FADANELLI, S. Leitura de textos especializados anglófonos no Ensino Técnico: idiossincrasias terminológicas e pedagógicas sob a perspectiva da Linguística de *Corpus*. **LínguaTec**, v. 1, n. 1, 2016, p. 1-28.

PALTRIDGE, B.; STARFIELD, S. **Thesis and dissertation writing in a second language**. London: Routledge, 2007.

ORENHA-OTTAIANO, A. Semelhanças e diferenças entre colocações e colocações especializadas. In: ORTIZ-ALVAREZ, M. L. (Org.). **Tendências atuais na pesquisa descritiva e aplicada em fraseologia e paremiologia**. 1 ed. Campinas: Pontes, 2012, p. 147-163.

PAIVA, V. L. M. O; PAGANO, A. S. English in Brazil with an outlook on its function as a language of science. In: AMMON, U. (Ed.). **The dominance of English as a language of science**. Mouton de Gruyter, 2001, p. 425-445. <https://doi.org/10.1515/9783110869484.425>

SARDINHA, T. B. **Linguística de Corpus**. São Paulo: Manole, 2004.

SARDINHA, T. B. **Linguística de Corpus**: Histórico e Problemática. **D.E.L.T.A.**, v. 16, n. 2, p. 323-367, 2000.

SARDINHA, T. B. **Pesquisa em Linguística de Corpus com WordSmith Tools**. Campinas: Mercado de Letras, 2009.

SCHMID, H. Probabilistic part-of-speech tagging using decision trees. **Proceedings of the International Conference on New Methods in Language Processing**. Manchester, United Kingdom, p. 44-49, 1994.

SANTOS, A. Análise de colocações adverbiais em inglês para negócios. In: **Corpora no Ensino de Línguas Estrangeiras**. VIANA, V.; TAGNIN, S. E. O. São Paulo: Hub Editorial, 2010, p. 97-136.

SWALES, J. M. **Genre Analysis** - English in Academic and Research Settings. Cambridge University Press, 1990.

SWALES, J. M.; FEAKE, C. B. **Academic Writing for Graduate Students: Essential Skills and Tasks**. 2nd edition. Ann Arbor, MI: University of Michigan Press, 2004.

TAGNIN, S. E. O. **O jeito que a gente diz** - combinações consagradas em inglês e português. Barueri/SP: Disal, 2013.

VIEIRA, R.; LOPES, L. Processamento de Linguagem Natural e o tratamento computacional de linguagens científicas. In: PERNA, C. B. L.; DELGADO, H. O. K.; FINATTO, M. J. B. (Org.). **Linguagens especializadas em corpora**: modos de dizer e interfaces de pesquisa. Porto Alegre: EDIPUCRS, 2010, p. 128-151.

Artigo recebido em: 13.12.2016

Artigo aprovado em: 06.04.2017