

LEVANTAMENTO DE PALAVRAS-CHAVE DA SEÇÃO DE COMPREENSÃO ORAL DO *TOEFL*: O USO DE *CORPORA* NA ÁREA DE INGLÊS PARA FINS ACADÊMICOS

*Jéssica Laira de Araujo Esgoti Uliana**

*Luciano Franco da Silva***

*Mileni Bertholini Ferreira****

RESUMO

O presente artigo teve como objetivo analisar um pequeno *corpus* compilado a partir da seção de compreensão oral do *TOEFL - Test of English as a Foreign Language*, e, posteriormente, comparar a frequência e chavicidade das palavras com o *corpus* MICASE – *Michigan Corpus of Academic Spoken English*, e o BASE – *British Academic Spoken English*. Dessa maneira, o artigo visou elencar uma lista das palavras-chave que poderão servir de base para criação de material didático preparatório para o teste. O estudo tem como referencial teórico as discussões sobre inglês para fins acadêmicos – IFA (ROBINSON, 1991; JORDAN, 1997 e GOH, 2012); testagem (MCNAMARA, 2009, 2000; FULCHER, 1999 e DOUGLAS, 2005, 2000) e uso de *corpora* no ensino de língua estrangeira – LE (FLOWERDEW, 2001; MCENERY; XIAO, 2011). Os resultados indicam uma lista de palavras relativamente mais frequentes em palestras acadêmicas da área de Ciências Biológicas, desta forma, as mesmas servem de possíveis recursos para o ensino de IFA.

Palavras-chave: *Corpora*; Compreensão Oral; IFA; Testagem; *TOEFL*.

ABSTRACT

The present paper aimed to analyze a small *corpus* compiled from the oral comprehension section of the *TOEFL - Test of English as a Foreign Language*, and then compare the word frequency and keyness with the MICASE *corpus - Michigan Corpus of Academic Spoken English*, and BASE - *British Academic Spoken English*. Thus, the article aimed to present a list of keywords that could serve as a basis for the creation of teaching material for the test. The study has as theoretical reference the discussions about English for Academic Purposes - EAP (ROBINSON, 1991; JORDAN, 1997 and GOH, 2012); And the use of *corpora* in foreign language teaching – EFL (FLOWERDEW, 2001; MCENERY, XIAO, 2011). The results indicate a word list relatively more frequent in Biological em Health Science academic lectures, this way, they serve as possible resources for the teaching of EAP.

Keywords: *Corpora*; Oral Comprehension; EAP; Testing; *TOEFL*.

* Mestre em Estudos Linguísticos pela UNESP, professora de Inglês e Português na ETEC de Fernandópolis. UNESP – Ibilce.

** Mestrando em Linguística Aplicada na UNESP – Ibilce.

*** Mestranda em Linguística Aplicada na UNESP – Ibilce..

INTRODUÇÃO

É notório que a globalização abrange vários setores, dentre eles o econômico, o tecnológico e também o educacional. Devido a esse desenvolvimento social, nota-se o grande crescimento do interesse pela aprendizagem da língua inglesa (LI). Em razão da influência internacional em nosso país, o conhecimento de LI se impõe como requisito essencial para uma ascensão profissional e acadêmica; sendo assim, faz-se necessário comprovar tal conhecimento por meio de avaliações e testes.

Este contexto é descrito por McNamara (2000, p. 4): “primeiramente, os testes de linguagem desempenham um papel poderoso na vida de muitas pessoas, agindo como porta de entrada em importantes momentos de transição na educação, no emprego e na mudança de um país para outro¹”. Ainda na mesma linha de pensamento, Martins (2013, p. 19), afirma que “avaliar é uma atividade que está sempre presente em nossa prática diária”. Assim, estamos a todo momento avaliando e sendo avaliados. Na área de ensino de LI, em geral, a competência linguística do aluno é avaliada por meio de exames de proficiência, e o resultado desses exames pode ter um peso muito grande na hora de tomadas de decisões, podendo proporcionar ou fazer perder oportunidades de ascensão profissional ou acadêmica dos alunos.

Em relação à avaliação de inglês para fins acadêmicos – IFA – existem diversos testes que visam atestar a proficiência de seu candidato. Os mais conhecidos são: *TKT - Teacher Knowledge Test*; *TEAP – Test of English for Academic Purposes*; *IELTS – International English Language Testing System*; *TOEFL - Test of English as Foreign Language*.

Dentre os exames mencionados, o *TOEFL* é o que mais se destaca, motivo pelo qual o presente artigo tem como objetivo analisar a seção de compreensão oral (CO) do teste, devido ao fato de a CO se apresentar como uma habilidade desafiadora na área de ensino-aprendizagem, por envolver tantos os processos linguísticos, como também crenças, conhecimento de mundo, habilidades metacognitivas, dentre outros fatores, sendo estes não totalmente controlados pelo professor (GOH, 2012; GONDIM, 2008; FLOWERDEW, 1994). Tal afirmação é tida como justificativa para o presente artigo, pois, pelo fato de a CO ser uma habilidade tão complexa quanto qualquer outra, o estudo poderá auxiliar no processo de preparação do candidato para esse teste de proficiência acadêmica.

Para a realização do estudo, foram selecionados os áudios que representavam palestras acadêmicas da área de Ciências Biológicas em simulado do teste *TOEFL*; posteriormente, esses dados foram comparados com o *corpus MICASE (Michigan Corpus of Academic Spoken English)*, obtendo, assim, listas dos vocábulos mais recorrentes que podem servir de apoio ao professor para criação de materiais preparatórios para essa seção do teste.

Sendo assim, o propósito do estudo baseia-se na afirmação de O’Keeffe, McCarthy e Carter (2007) de que os dados obtidos por *corpora* são exemplos da língua usados em contextos autênticos, podendo esses seres de grande valia real para a preparação do aluno para esse contexto específico. O artigo foi dividido em: *i*) referencial teórico, *ii*) metodologia, *iii*) resultados e discussões e *iv*) considerações finais.

1 Tradução dos próprios autores: “*First, language tests play a powerful role in many people’s lives, acting as gateways at important transitional moments in education, in employment, and in moving from one country to another.*”

1 REFERENCIAL TEÓRICO

1.1 Definindo os conceitos de avaliação e testagem

Visto que o ato de avaliar está presente a todo o momento em nossa vida e em qualquer área gera grande preocupação, o termo “avaliação” é amplamente discutido nos segmentos escolares e acadêmicos. Em primeiro lugar, podemos afirmar que essa palavra, em muitos casos, é utilizada como um termo guarda-chuva, representando tanto os métodos de testagem (*testing*) como os avaliativos (*assessment*) (CLAPHAM, 2000). Segundo o mesmo autor, a testagem seria constituída pelas construções e administrações de testes formais ou padronizado e de larga escala, como o TOEFL. Já os métodos avaliativos são aqueles mais informais, relacionados ao contexto escolar. Sendo assim, a avaliação, em qualquer dos seus sentidos, está intrinsecamente relacionada com o processo de ensino-aprendizagem.

Todavia, Luckesi (2008, p. 21) apresenta uma crítica ao modo como a avaliação é utilizada na maioria dos contextos educacionais, afirmando que “a nossa prática educativa se pauta por uma ‘pedagogia do exame’”, ou seja, a avaliação passou então a ser interpretada somente como um instrumento somativo focalizando apenas no produto de aprendizagem. Segundo Haydt (2000), a avaliação somativa é aquela que tem como função classificar os alunos ao final da unidade, semestre ou ano letivo, segundo níveis de aproveitamento, classificando o aluno muitas vezes como aprovado ou reprovado. Alguns exemplos de avaliação somativa são as avaliações classificatórias, de ranqueamento e psicométrica.

Em oposição à ideia de avaliação somativa, muitos autores defendem a utilização de um modelo formativo de avaliação. De acordo com Scaramucci (1999, p. 118), “Hoje fala-se em avaliação de processo, contínua, não apenas somativa mas também formativa; processo dinâmico, interativo e negociável, que extrapola o uso de testes e provas; avaliação individualizada, que leve em conta aspectos afetivos”. Portanto, pode-se afirmar que o processo de ensino-aprendizagem está ligado tanto às avaliações somativas como às formativas.

Em relação à testagem, Brown (2001, p. 384) a define da seguinte maneira: “um método de mensurar a habilidade ou conhecimento de uma pessoa em um determinado domínio”. O autor ainda complementa afirmando que um teste é um conjunto de técnicas, procedimentos e itens que constituem um instrumento que requer alguma atividade ou desempenho por parte de outra pessoa.

Em relação aos testes no campo do ensino de línguas, McNamara (2009) afirma que eles estão cada vez mais complexos e que a sua qualidade técnica e científica se apresenta como fonte do seu poder perante a sociedade. Entre os testes mais usados, podemos citar os já mencionados: *TOEFL* e *IELTS*, assim como o *TOEIC (Test of English for International Communication)*, testes de proficiência utilizados como prova de habilidade linguística perante a sociedade e forma de ascensão acadêmica e profissional (CLAPHAM, 2000; MCNAMARA, 2009; 2000).

1.2 Avaliação em IFE

Assim como o ensino em IFE é diferente do ensino geral da língua inglesa, por objetivar as necessidades de cada aprendiz, a avaliação também difere em alguns aspectos. Os conceitos básicos de avaliação, tais como confiabilidade, validade e impacto do teste também são encontrados em um

Inglês para Fins Acadêmicos e *corpora*

A pesquisa com *corpus* tem ganhado grande destaque na área do IFA, em especial no que se refere à descrição e à análise do discurso acadêmico (FLOWERDEW, J., 1994; FLOWERDEW, L., 2001; MCENERY e XIAO, 2011). Segundo Flowerdew (1994), esse tipo de pesquisa pode direcionar professores e pesquisadores sobre as características linguísticas e discursivas com que os alunos precisam estar familiarizados para otimizar a compreensão de palestras e, portanto, quais os conteúdos que devem ser incorporados a um curso de IFA.

É válido destacar aqui alguns *corpora* eletrônicos, tais como os *corpora* BAWE³ (*British Academic Written English Corpus*), o BASE⁴ (*British Academic Spoken English*) e o MICASE⁵ (*Michigan Corpus of Academic Speech*). *Corpora on-line*, como esses, são excelentes recursos para professores e pesquisadores que queiram obter informações mais detalhadas sobre textos acadêmicos.

Dentre as diversas utilizações de *corpora* no ensino, Charles (2012, p. 139-140) destaca três usos de *corpora* dentro do IFA. São eles: a *i) identificação de repetições de padrões linguísticos* dentro de uma quantidade massiva de dados, revelando assim, evidências objetivas, e não mais subjetivas, na análise linguística; *ii) descrição de características específicas do discurso acadêmico*, assim como a fraseologia de diferentes disciplinas e gêneros acadêmicos, e, por último, *iii) corpora* são usados na *compilação de listas de palavras acadêmicas*⁶.

Complementando os usos de *corpora* mencionados acima, Flowerdew (2001) apresenta o uso de *corpora* pequenos na produção de material voltados ao IFA, visto que, para o contexto acadêmico, eles são mais especializados em termos de tópico e gênero.

Em relação ao tamanho ideal de um *corpus*, existem discussões entre linguistas, tradutores e lexicólogos e, até hoje, ainda não chegaram a um consenso (VIANA, 2010). Dentro dessa discussão, alguns pesquisadores, como Sinclair (2004), defendem a ideia de que quanto maior melhor, por outro lado, Tribble (2001) argumenta que apesar de a ideia de grandes *corpora* ser válida para determinados tipos de pesquisa, isso não necessariamente é válido para professores e aprendizes de IFA, uma vez que grandes *corpora* demonstram uma limitada relevância de dados para quem tem interesse na área de línguas para fins específicos.

Defendendo o ponto de vista de Tribble sobre o uso de *corpora* pequenos, Bowker e Pearson (2002, p. 46) explicam que

É muito importante, no entanto, não assumir que o maior é sempre melhor. É possível descobrir informações mais úteis a partir de um *corpus* pequeno, porém bem planejado, do que em um *corpus* grande, mas que não atende às suas necessidades⁷.

3 Disponível em: <http://www2.warwick.ac.uk/fac/soc/al/research/collections/bawe/>

4 Disponível em: <http://www2.warwick.ac.uk/fac/soc/al/research/collections/base/>

5 Disponível em: <http://quod.lib.umich.edu/cgi/c/corpus/corpus?c=micase;page=simple>

6 Maiores informações: <http://www.victoria.ac.nz/lals/resources/academicwordlist/>

7 Nossa tradução de *It is very important, however, not to assume that bigger is always better. You may find that you can get more useful information from a corpus that is small but well designed than from one that is large but is not customized to meet your needs.*

Dessa forma, conforme indica Viana (2010, p. 29), “a adequação de um *corpus* deve prevalecer sobre questões acerca de seu tamanho. Dessa forma, garante-se que os dados sejam representativos do uso da linguagem que se deseja investigar”.

2 METODOLOGIA

O *corpus* de estudo para a pesquisa foi compilado a partir de 14 palestras acadêmicas retiradas da seção de *listening comprehension* do TOEFL, todas pertencentes à área de Ciências Biológicas. Em uma primeira análise, foi possível obter as seguintes informações: quantidade de itens (*types*), número de formas (*tokens*) e a densidade lexical do *corpus*; a tabela 1 apresenta as informações mencionadas anteriormente.

Tabela 1 – Características lexicais do *corpus* de estudo

No. total de itens	11,670
No. total de formas	2,692
Densidade lexical	23%

É importante salientar que, quanto maior a densidade lexical, maior é a variedade no uso do léxico, ou seja, maior a individualidade das palavras nos os áudios utilizados no teste. A partir disso, buscou-se extrair a lista de palavras mais frequentes dentro do *corpus*, permitindo, assim, a verificação dos itens lexicais mais utilizados.

Quadro 4 – Lista de palavras lexicais mais recorrentes do *corpus* de estudo

Item	Frequência	Item	Frequência
1. food	52	11. seeds	21
2. other	52	12. percent	20
3. species	51	13. plant	20
4. birds	44	14. nuts	19
5. animals	27	15. bees	18
6. right	26	16. water	18
7. new	24	17. both	18
8. now	23	18. back	17
9. scientists	22	19. into	16
10. islands	21	20. way	16

Fonte: elaborado pelos autores

Em último momento, partimos para a identificação das palavras-chave, ou seja, palavras estatisticamente mais significantes no *corpus* de estudo, quando comparadas a um *corpus* de referência. Para isso, verificou-se a chavicidade (*keyness*⁸) das palavras do *corpus* de estudo, quando comparadas com a seção de ciências biológicas do *corpus* BASE, visto que este é mais de cinco vezes maior do que nosso *corpus* de estudo, segundo recomendado por Berber Sardinha (2009).

8 O termo chavicidade é o resultado de um procedimento estatístico que levanta o quão importante cada palavra-chave é para o *corpus* de pesquisa em relação ao de referência. Quanto maior o valor apresentado, maior a relevância da palavra em questão (VIANA, 2010).

É importante lembrar que, nos estudos com *corpora*, palavras-chave não são o mesmo que palavras “importantes”, pois os programas de análise utilizam um critério estatístico, quantitativo, para identificar as mesmas (BERBER SARDINHA, 2009), abaixo segue as 20 palavras-chave estatisticamente mais significantes em nosso *corpus* de estudo.

Quadro 5 – Palavras-chave encontradas no corpus de estudo em comparação com o BASE

	<i>keyness</i>		<i>keyness</i>
1. food	198.461	11. bees	68.698
2. spices	178.255	12. percent	68.612
3. birds	158.659	13. scientists	56.795
4. island	80.148	14. water	56.337
5. seeds	80.148	15. dance	53.432
6. animals	76.915	16. paradise	53.432
7. eat	76.331	17. autoclave	49.615
8. plant	76.331	18. fruit	45.799
9. nuts	72.515	19. fish	42.731
10. bees	68.698	20. energy	41.982

Fonte: Elaborado pelos autores

Uma outra análise importante, foi a definição da especificidade das palavras-chave encontradas, para isso, utilizou-se o *corpus* acadêmico MICASE para a análise estatística das frequências das palavras-chave na área de Ciências Biológicas, em relação às outras áreas do conhecimento acadêmico.

Quadro 6. Frequência das palavras dentro do corpus MICASE.

	Ciências Biológicas	Humanidades	Exatas e Tecnológicas	Ciências Sociais	Outros	Total	% na área de C. Biológicas
<i>food</i>	94	31	20	38	12	195	48,2
<i>species</i>	545	4	39	19	1	608	89,6
<i>birds</i>	139	3	1	3	0	146	95,2
<i>island</i>	8	9	15	12	0	44	18,1
<i>seeds</i>	28	7	5	1	0	41	68,2
<i>animals</i>	115	28	4	12	0	159	72,3
<i>nuts</i>	2	2	2	7	0	13	15,3
<i>percent</i>	124	36	165	218	24	567	21,8
<i>plant</i>	152	4	8	13	4	181	83,9
<i>black</i>	33	62	41	191	5	332	9,9
<i>scientists</i>	21	2	2	12	11	48	43,7
<i>water</i>	155	54	275	30	21	535	28,9
<i>dance</i>	4	11	2	13	8	38	10,5
<i>hole</i>	27	11	16	4	9	67	40,2
<i>paradise</i>	1	0	0	0	0	1	100
<i>autoclave</i>	0	0	0	0	0	0	0
<i>energy</i>	131	8	225	12	6	382	34,2
<i>fruit</i>	38	11	8	3	0	60	63,3
<i>sun</i>	36	19	32	5	0	92	39,1
<i>fish</i>	188	29	15	4	6	242	77,6

Fonte: Elaborado pelos autores

Analisando o quadro anterior, é possível perceber que a maioria das palavras-chave encontradas na seção de CO do *TOEFL*, quando comparadas ao *corpus MICASE*, são mais frequentes dentro da área de Ciências Biológicas, podendo as mesmas guiarem o professor durante a preparação das suas aulas.

CONSIDERAÇÕES FINAIS

O intuito do estudo apresentado foi compilar uma lista de palavras-chave da área de Ciências Biológicas, presentes na seção de CO do teste *TOEFL*, para utilização em sala de aula. Visto que uma das maiores dificuldades que os alunos enfrentam na seção de CO são os vocábulos apresentados, neste estudo focamos na área de Ciências Biológicas, o que também representa uma dificuldade para o professor na preparação de materiais para o referido teste, já que ele não tem como saber em quais vocábulos deverá priorizar durante as aulas. Dessa forma, a linguística de *corpus* se apresenta como um excelente recurso teórico-metodológico para o levantamento lexical, e o levantamento de palavras-chave se mostra como um excelente guia, tanto para o professor quanto para o aluno.

Limitações do estudo

O estudo realizado apresentou limitações quanto à sua população e amostra. A população definida para este estudo, 14 palestras acadêmicas do TOEFL pertencentes à área de Ciências Biológicas, se apresenta como limitante para maiores sugestões de ensino em IFE/IFA. Como apresentado no decorrer deste artigo, o *corpus* de estudo conta apenas com 11.670 palavras. Todavia, considerando os resultados encontrados, o fato de o *corpus* de estudo ser pequeno, não diminui a sua adequação ao estudo.

REFERÊNCIAS

- BERBER SARDINHA, T. **Pesquisa em linguística de corpus com WordSmith tools**. Campinas: Mercado de Letras, 2009.
- BOWKER, L.; PEARSON, J. **Working with specialized language: a practical guide to using corpora**. Routledge. 2002
- BROWN, H. D. Language Assessment I: Basic concepts in test development. In: **Teaching by principles: An interactive approach to language pedagogy**. 2^a ed. New York: Longman, 2001. Chapter 21, p. 384 – 400.
- CHARLES, M. English for Academic Purpose. In: PALTRIDGE, Brian; STARFIELD, Sue. (Org) **The Handbook of English for Specific Purposes**. Chichester, UK. John Wiley & Sons, Ltd. 2012. p. 137-154.
- CLAPHAM, C. **Assessment and testing**. Annual Review of Applied Linguistics. 20, 2000.
- DAVIES, A. **The logic of testing languages for specific purposes**. Language Testing18: 133-48. 2001

DOUGLAS, D. **Assessing Language for Specific Purposes**. Cambridge: Cambridge University Press. 2000

_____. Testing language for specific purposes. In E. Hinkel (ed), **Handbook of Research in Second Language Teaching and Learning**. 857-68. Mahwah, NJ: Lawrence Erlbaum. 2005.

FLOWERDEW, J. Research of relevance to second language lecture comprehension – an overview. In: FLOWERDEW, John. (Org.) **Academic listening: Research perspectives**. Cambridge University Press, 1994. p. 7-30, 1994.

FLOWERDEW, L. The exploitation of small learner corpora in EAP materials design. In: GHADESSY, Mohsen; HENRY, Alex; ROSEBERRY, Robert L. (Ed.). **Small corpus studies and ELT: theory and practice**. John Benjamins Publishing, Amsterdam, p. 363-379, 2001.

FULCHER, G. **Assessment in English for academic purposes: putting content validity in its place**. Applied Linguistics 20:221-36. 1999

GOH, C.C.M. Esp and Listening. IN: PALTRIDGE, Brian; STARFIELD, SUE. (Org) **The Handbook of English for Specific Purposes**. Chichester, UK. John Wiley & Sons, Ltd. 2012. p. 55-76.

GONDIM, M. L. S. F. **Um novo olhar sobre a compreensão oral: os mecanismos subjacentes ao ensino da compreensão oral segundo o enfoque da abordagem comunicativa**. 2008.

HAYDT, R. C. **Avaliação do processo ensino-aprendizagem**. São Paulo: Ática, 2000.

JORDAN, R. R. **English for Academic Purposes: a guide and resource for book for teachers**. Cambridge: Cambridge University Press. 1997.

LUCKESI, C. C. **Avaliação da aprendizagem escolar: estudo e proposições/ Cipriano Carlos Luckesi**. - 19ª ed. São Paulo: Cortez, 2008.

MARTINS, T. H. B. **Análise e validação de um teste de classificação em inglês: um estudo de caso em cursos superiores em tecnologia**. Tese de Doutorado. Universidade de São Paulo.

MCENERY, T.; XIAO, R. What corpora can offer in language teaching and learning. **Handbook of research in second language teaching and learning**, v. 2, p. 364-380, 2011.

McNAMARA, T. **Testing, testing... What is a language test?** Language testing. Oxford: oxford university Press, 2000.

_____. Principles of testing and assessment. KNAPP, K; SEIDLHOFER, B (EDS.) **Handbook of foreign language communication and learning**. Berlin / New. York: Mouton de Gruyter, 2009.

O'KEEFFE, Anne; MCCARTHY, Michael; CARTER, Ronald. **From corpus to classroom: Language use and language teaching**. Cambridge University Press, 2007.

O'SULLIVAN, Barry. **Issues in testing business English: the revision of the Cambridge Business English Certificates**. Cambridge University Press, 2006.

PILEGGI, M. G. S. **Tarefas integradas nos exames de proficiência CELPE-BRAS e TOEFL iBT**. Campinas, SP: [s.n.], 2015. Dissertação de mestrado – Universidade Estadual de Campinas, Instituto de Estudos da Linguagem.

ROBINSON, P. **ESP Today: A Practitioner's Guide**. New York: Prentice Hall. 1991.

SCARAMUCCI, M.R.V. **Avaliação: mecanismo propulsor de mudanças no ensino/aprendizagem de língua estrangeiras**. Revista Contexturas. São Paulo, APLIESP, 1999.

SCOTT, M. **Wordsmith tools**. Oxford: Oxford University Press, 1999.

THIGHE, DAVID. Cambridge ESOL and tests of English for Specific Purposes. **Research Notes**, v. 27, p. 2-4, 2007.

TRIBBLE, C. Small corpora and teaching writing. In: GHADDESSY, M.; HENRY, A.; ROSEBERRY, R. L. (Ed.). **Small corpus studies and ELT: theory and practice**. John Benjamins, Amsterdam, 2001. p. 381 - 408.

SINCLAIR, J. M. **Trust the text: language, corpus, and discourse**. London: Routledge, 2004.

VIANA, V. Linguística de *corpus*: conceitos, técnicas & análises. In: VIANA, V.; TAGNIN, S. E. O. TAGNIN (Org.). **Corpora no ensino de línguas estrangeiras**. São Paulo: Hub Editorial, 2010. p. 25 – 96.