

医学学术英语两栖词汇研究

——基于语料库的词表建立和教学策略

何晴

(福建医科大学 文理艺术学院, 福建 福州 350122)

摘要:两栖词汇是学生学习医学学术英语的难点,教师由于缺乏相应的词表作为参照,也难以将两栖词汇全面、系统地纳入医学生的英语教学。研究旨在通过建立词表为师生提供参照,提高该类词汇教与学的效率。采用在线语料分析工具将 MedDEAP 语料库和 Crown 语料库进行对比得出医学学术英语核心词汇,再从中筛选出两栖词汇。研究发现,核心技术性词汇覆盖 MedDEAP 全库的 2.91%,其中 75%是两栖词汇,由亚技术性词汇和半技术性词汇组成,其搭配呈现出了突出的语域特征。医学学术英语两栖词汇教学中应优先处理核心两栖词表中的词汇,重视目标词在更大语境范围内的共现词,借助语料库索引行让学生在数据驱动下积极主动地探究两栖词汇。

关键词:语料库;医学学术英语;两栖词汇;词表;教学策略

中图分类号:H319.3;R-4

文献标志码:A

文章编号:1009-4784(2022)01-0065-06

医学英语作为专门用途英语(English for special purpose, ESP)中的特殊文体,词汇、语法和语篇方面都有其特征,在医学学术文体中,词汇特征尤为明显。该文体中通用词汇在日常生活中使用频率高,属于语言的共核部分,易于理解^[1];纯医学词汇是仅在医学领域使用的术语,一词一义,也不含混。而两栖词既是普通英语词汇,又具有医学含义,一词多义,是造成学生学习的难点,被认为是医学学术英语词汇教学中应该优先考虑的^[2]。但教师由于缺乏相应的词表作为参照,难以将两栖词汇全面、系统地纳入医学生的英语教学。本研究通过语料库方法建立医学学术英语两栖词表,以为师生提供参照,提高该类词汇教与学的效率。

目前有关医学学术英语中两栖词汇的研究,多聚焦于认知过程或翻译技巧,从经验直觉的角度出发,经过归纳整理列举实例,能够穷尽的例证有限,且只见其木不见其林。而语料库则提供了真实语言大数据,能由检索项推及更大的语境,揭示倾向性规律,且结果可以验证。虽不乏研究运用语料库方法研究医学英语高频词汇^[3-4]或医学学术英语高频词汇^[5],但有关医学学术英语两栖词汇的现有研

究主要是质性研究,少有依托语料库进行的定性定量相结合的研究。

目前,国内医学英语词汇教学偏重医学领域独有的术语,从词源和构词法等方面进行系统讲授,学生通过了解词汇文化出处、分析词根和词缀推知这些术语的意思。而虽然两栖词汇在医学文本中占据重要地位,教学中却并没有整合零散分布在文本中的此类词汇,甚至没有明确将其和普通词汇进行区分。因此,学生往往忽略对该类词汇有意识的学习^[2]。

本研究先明确两栖词汇的界定方式,以在线语料库为主要工具,配合定性分析,建立医学学术英语两栖词表,分析两栖词汇在医学学术英语中的分布特点、搭配习惯。同时,探讨如何借助该词表以及语料库工具的辅助,改进医学英语词汇教学,以期帮助学生更有针对性地扫除两栖词汇的障碍,从而提高学习效率。

一、医学两栖词汇的界定

对医学词汇的划分,较多研究采用 Nation 对书面学术文本中词汇的分类^[6]:高频词(high-frequen-

cy words)广泛使用于日常生活;学术性词汇(academic words)广泛出现于各学科学术性文本;技术性词汇(technical words)与学科专业密切相关;低频词(low-frequency words)较常使用于学术文本但频率也很低。但是该分类未细分技术性词汇,未体现纯术语(纯医学词汇)和半术语(两栖词汇)的区别^[2]。且在有些研究中,学术性词汇被冠以“半技术性词汇”这一名称^[7],易造成学术性词汇和技术性词汇的混淆。定义不清给后续研究的开展带来困难,若无明确界定则无法建立词表。

本研究以 Menon 和 Mukundan 对医学技术性词汇的详细分类^[8]为理论基础,筛选语料库中的两栖词汇。Smith 和 Davis 对该分类进行了解读^[9]:纯技术性词汇(highly technical words)在其他领域少有使用,如源自希腊语和拉丁语的医学词汇;亚技术性词汇(sub-technical words)各学科都常用,意义随学科变化有所不同,但和通用义十分相近,易从通用义推知,对语境依赖不大;半技术性词汇(semi-technical words)的意义和特定学科相关,依赖语境,和通用义联系较弱,较难从通用义中推知词义,甚至可独立于通用义存在。

在此分类指导下,两栖词汇实为亚技术性词汇和半技术性词汇的总和。国内学者也有类似的归纳,将两栖词汇分为引申词汇(技术含义与通用义有直接关系)和独立词汇(技术含义与普通含义无明显联系)^[10]。

二、研究方法

(一)工具

此前运用语料库的医学英语词汇研究,多基于自建的小型语料库,使用 AntConc、Wordsmith 等第三代语料分析工具。每个研究采集的文本素材均不同,需要重新清洁、标注,效率受限且欠缺可重复性。本研究则采用北外 CQPweb 平台上集成的 MedDEAP 和 Crown 语料库,既能快速得到检索项并进行分析,也能保持语料来源的相对稳定。CQPweb 属于第四代语料库工具,即在线语料库分析工具,将语料库与分析工具合二为一,检索响应速度和操作便捷程度均远高于第三代软件^[11]。

MedDEAP 是北外中国外语与教育研究中心学术英语数据库(database of English for academic purposes, DEAP)项目的临床医学部分,开放在线使用的 V1.0 版本容量 1 216 231 词,收集了“临床

医学”(学科代号 1002)下 18 个二级学科各自领域影响因子前 10 名之内的 3 至 5 本国际期刊优秀论文全文及摘要^[12]。

Crown 是于 2012 年建成的通用美语语料库,由北外中国外语教育研究中心组织国内 107 所高校 150 多位外语教师及研究生建立,取样标准及语料规模与 Brown 语料库完全一致,且比 Brown 更具时效性,容量 1 026 226 词^[13]。

(二)分析方法

首先,将 MedDEAP 作为“观察语料库”(observed corpus),Crown 作为“参照语料库”(reference corpus),设置最低频数 5 次,比较得出 MedDEAP 中的正向核心词汇。其次,剔除结果中的标点符号、功能词等和文体关系不大的成分。再次,界定列表中的高频词汇、学术词汇和技术性词汇,并区分技术性词汇中的亚技术性词汇、半技术性词汇和纯技术性词汇。剔除高频词汇、学术词汇及纯技术性词汇,剩余亚技术性词汇和半技术性词汇,即两栖词汇。其中,高频词汇的界定参照《通用英语词表》(a general service list of English words, GSL)中的 2 000 个词族^[14],学术词汇的界定参照《学术英语词表》(academic word list, AWL)中的 570 个词族^[15]和《学术词表》(academic vocabulary list, AVL)^[16],3 类技术性词汇的界定参照上文中 Menon 和 Mukundan 的分类方法^[8],在《牛津词典》和《道兰医学图解词典》的辅助下进行定性分析判断。最后,锁定高频两栖词汇后,形成词表并对照含义(仅取 MedDEAP 中常见的医学含义),对其整体分布和搭配特点等进行描述分析,并对教学提出建议。

三、两栖词汇在医学学术文本中的分布和特点

(一)两栖词汇覆盖率高

经过观察库和参照库的比较,剔除标点符号、虚词等之后,本研究得到的核心词表共 1 053 词,各类词汇分布情况。其中亚技术性词汇的总类符数为 75,总形符数为 13 215,半技术性词汇的总类符数为 67,总形符数为 13 362,二者覆盖率接近,四舍五入后均为 1.09%。二者合并即两栖词汇,共出现 26 577 次,覆盖率为 2.18%。Coxhead 的研究发现

技术性词汇通常构成学术性文章中约 5% 的行文文字^[15]，而本研究的数据显示仅仅 MedDEAP 核心词表中的技术性词汇就覆盖了全库的 2.91%，其中 75% 是两栖词汇。若在教学中能处理好这一部分

词汇，将对技术性词汇的理解和掌握大有裨益。相较于出现频率较低的纯技术性词汇，对两栖词汇通用含义和医学含义的区分能力更能反映医学学术英语学习效果的优劣(表 1)。

表 1 核心词表中各类词汇分布

词汇类型	高频词汇	学术词汇	技术性词汇		
			两栖词汇		纯技术性词汇
			亚技术性词汇	半技术性词汇	
类符数/个	336	505	75	67	70
形符数/个	112 621	85 284	13 215	13 362	8 853
覆盖率/%	9.26	7.01	1.09	1.09	0.73

结果显示，核心程度排序第一的两栖词汇是“treatment”(对数似然率比 1 296.05)，排序最后一位的是“compression”(对数似然率比 16.06)。两栖词表涉及 106 个词族，其中亚技术性词汇涉及 55 个词族，半技术性词汇涉及 51 个词族。同一词族中不同词形在两栖词表中的排序有所不同，意味着它们的重要程度有差异，没有进行词形还原就是为了保留不同词形在 MedDEAP 中的重要性信息。如“progression”和“progressive”分别排在第 14 位和 50 位，说明“progression”在医学学术英语文体中比“progressive”拥有更核心的地位。为便于检索，词表采用词族归类的方式呈现，但应标注出每个词形的位次排序。同词族的几个词若不同义则分列，若同义则合并呈现释义。

研究发现以下几个词只覆盖 1~2 个子库，代表性稍弱。“deliveries”只在妇产科学子库和麻醉学

子库中出现，“SECT”只在影像医学与核医学子库中出现，“lens”只在眼科子库和麻醉学子库中出现。但同词族的其他词则不存在代表性问题，如“delivery”覆盖 15 个子库，“sections”覆盖 14 个子库。

(二)中段核心两栖词汇表意丰富

核心词表的 1 053 词被等分为三部分，排序 1—351 位的词为第一组，排序 352—702 位的词为第二组，排序 703—1 053 位的词为第三组，两栖词汇的分段分布情况见表 2。

从第一组到第三组，两栖词汇的形符数逐步降低，其中亚技术性词汇和半技术性词汇的形符数均逐步降低(图 1)；但类符数则呈现出中段较其他两段多的趋势(图 2)。前研究曾发现词表中的中段词通常是一些灵活、表意丰富的词^[17]，中段两栖词汇突出的类符数印证了其具备灵活表意的特征。

表 2 核心词表中两栖词汇分段分布

词汇类型	亚技术性词汇		半技术性词汇		两栖词汇	
	类符数/个	形符数/个	类符数/个	形符数/个	类符数/个	形符数/个
第一组	25	8 533	24	9 209	49	17 742
第二组	28	3 249	29	3 331	57	6 580
第三组	22	1 433	14	882	36	2 315

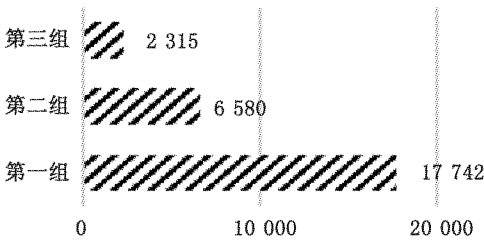


图 1 核心词表中两栖词汇形符数分布

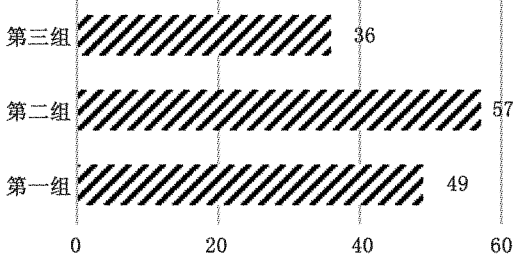


图 2 核心词表中两栖词汇类符数分布

(三)两栖词汇的搭配存在跨语域差异

通过搭配强度计算,关键性排名前 10 位的两栖词汇在医学学术英语文本中的搭配特点得以揭示。跨距设定为 ± 4 ,节点词至少覆盖 5 篇文章且频数至少 10 次,按对数似然比高低排列,去除虚词、标点以及覆盖文章篇数过少或总频率过少的结果后,搭配强度前三位的词见表 3。

在医学学术文本中,“treatment”是指治疗或对病人的处置,与之共现程度最高的几个词分别是“groups”“options”和“effect”,用于在探究治疗手段时描述实验过程或结果等,文体特征突出。“cells”和“cell”虽只是单复数的区别,但从搭配上,复数形式的 cells 通常与前一个词一起,指具体种类的细胞(endothelial cells, epithelial cells, progenitor cells),而单数形式的 cell 通常作后一个词的修饰词,指细胞相关的状态(cell death, cell lines)或指征(cell count)。risk factor (s)这一强搭配说明医

学学术文章重分析造成疾病等不良后果的因素,而 risk 在通用语料库 Crown 中则倾向和动词组成强搭配,如“reduce”“increase”“taking”等。“stroke”(中风)和“disorders”(紊乱)的强搭配词则反映出问题产生的主要原因,如“ischemic”“heart”“cardiovascular”表示中风多和心血管方面的问题有关,“psychiatric”“mood”和“mental”则说明 MedDEAP 中提到的紊乱主要指精神、情绪方面的紊乱。“function”主要指机体的功能,因此搭配的多是和身体结构相关的词汇(physical, mitochondrial 和 structure)。activity 除了指身体活动(physical activity)外,在很多情况下指活性(specific activity 和 activity levels)，“activation”的释义也与此相关,意为“使具有活性”,在跨距内常和“microglial/microglia”“pathway”共现。“tissue”的显著搭配词与其共同指称某种类型的人体组织(soft, adipose, normal)。

表 3 代表性两栖词汇前三位强搭配

排序	节点词	搭配强度第一位 (对数似然比)	搭配强度第二位 (对数似然比)	搭配强度第三位 (对数似然比)
1	treatment	groups (184.800)	options (142.114)	effect (104.268)
2	cells	endothelial (295.893)	epithelial (269.251)	progenitor (282.376)
3	risk	factors (1 081.135)	cardiovascular (429.107)	factor (424.063)
4	cell	death (277.810)	lines (257.851)	count (161.982)
5	stroke	ischemic (591.005)	heart (34.329)	cardiovascular (26.497)
6	disorders	psychiatric (650.860)	mood (233.647)	mental (118.418)
7	function	physical (177.051)	mitochondrial (144.491)	structure (137.665)
8	activity	physical (802.838)	specific (94.681)	levels (65.183)
9	tissue	soft (236.428)	adipose (207.330)	normal (110.994)
10	activation	microglial (159.928)	pathway (48.153)	microglia (46.453)

以上所列节点词的典型搭配在 MedDEAP 中的对数似然高于通用语料库 Crown,说明它们在医学英语中吸附共现词的能力比在通用英语中强得多,且前三位的显著搭配词均充分体现了医学学术文体的内容特点,进而证明了医学英语两栖词汇的用法具有跨语域的差异。

四、对两栖词汇教学的建议

(一)讲究重点和方法,提高学习效率

首先,医学词汇教学中应优先处理高频的核心词汇,而本研究发现核心词汇中的技术性词汇有

75%是两栖词汇,两栖词表能助力学生高效掌握绝大部分最重要的技术性词汇。优先处理核心词汇中的两栖词汇,对教师而言是把精力和时间用在合理的地方,对学生而言则可循序渐进地推进医学英语词汇学习^[3]。其次,医学学术两栖词汇的学习可分阶段开展,依据上文对医学两栖词汇的界定,区分各阶段侧重点:在基础英语学习阶段,应关注亚技术性词汇(医学含义与通用含义接近的词汇),可用联想法帮助记忆^[18],如从“case”的通用义“事件”联想到“患病的事件”即“病例”,适当介绍这些通用词的医学引申义,符合该阶段学生水平,同时也能衔接基础英语学习和专业英语学习;在专业英语学

习阶段,关注重点则应转移到半技术性词汇上,因其医学含义跟通用义相距较远,联想的难度较大,学生可借助本研究建立的词表有意识地记忆。

(二)从搭配强度入手,理解词汇用法

词汇学习中孤立地看待词有着显而易见的不足^[17],还应考虑词与词之间的结伴关系。Sinclair也提出了除了词本身,还应学习词汇的基本使用模式及搭配词^[19]。本研究则发现两栖词汇在医学英语中吸附共现词的能力比在通用英语中强得多,且其搭配存在跨语域差异。因此,在医学学术英语两栖词汇的学习中,更应重视搭配强度的概念。

教师可参考上文做法,运用语料库工具锁定目标词在更大语境范围内的共现词,通过比较目标词在医学英语语料库和通用英语语料库中强搭配词的区别,帮助学生更好地理解其用法,从而在产出时更准确。除此之外,两个词如果因为语义的重合而产生理解和运用的困惑,也可以通过考察它们的搭配词及其语义表达功能的差异予以消除^[20]。如两栖词汇“disruption”和“intervention”都有打断之意,通过医学英语语料库对比二者的强搭配词发现,前者常和身体上的部位一起出现,如“photoreceptor disruption”(光感受器缺失)和“central tendon disruption”(中央肌腱断裂),而后者常和“therapeutic”“exercise”等词共现,指治疗行为上的介入,由此可避免二者的混用。

(三)利用语料库索引行,激发学习主动性

在线语料分析工具不应仅仅供医学学术英语两栖词汇研究使用,还可辅助教学。本研究发现中段核心两栖词汇表意丰富,教师可从中选词并提取索引行交由学生观察。学生对目标词的搭配和用法进行假设和推断,再通过同伴讨论,最后归纳出合理的结论。这与交际教学中的解决问题、自我发现和互动讨论的做法相契合^[17]。为避免学习负荷过量,打击学习积极性,教师应从在线语料分析工具搜索出的完整索引行中进行抽样,建立微型文本^[17],筛选标准是其用法是否典型^[21]。教师甚至可根据医学生的具体专业在医学英语语料库的相应子库中筛选索引行,个性化的语料能够激发学生主动探究两栖词汇的兴趣。

五、小 结

本研究将语料库方法运用于领域类学术词表研究,对学术两栖词汇进行界定、筛选、分布和搭配

特点的描述,为医学生了解医学学术词汇和医学学术文体提供了有数据支持的新视角,也为医学生更高效地学习该文体中的两栖词汇提供了参考的范围。医学英语词汇教学中除了可参照词表确定教学重点,还可运用在线语料分析工具,从搭配入手辩词析义,利用丰富的语料数据驱动学生自主学习、合作学习,从而提高学生医学学术英语词汇的理解能力和产出能力。今后的研究可进一步探究该词表在医学生词汇学习中产生的实际效果。

参考文献:

- [1]王世杰,赵玉华,武永胜.基于语料库的医学英语词汇研究[学位论文].兰州:兰州大学出版社,2013.
- [2]谌谐婉,孙建国,陈敏,等.医学教育中学术英语词汇教学需求与难点分析[J].中华医学教育杂志,2017,37(4):552-556.
- [3]王连柱.基于语料库的高频医学英语词汇和词素的筛选与医学英语词汇教学[J].中华医学教育杂志,2011,31(1):95-97.
- [4]HSU W. Bridging the vocabulary gap for EFL medical undergraduates: the establishment of a medical word list[J]. Language Teaching Research, 2013, 17(4): 454-484.
- [5]LEI L, LIU D. A new medical academic word list: a corpus-based study with enhanced methodology[J]. Journal of English for Academic Purposes, 2016, 22: 42-53.
- [6]NATION P. Learning vocabulary in another language[M]. Cambridge: Cambridge University Press, 2001.
- [7]陈琦,高云.学术英语中的半技术性词汇[J].外语教学, 2010, 31(6): 42-46.
- [8]MENON S, MUKUNDAN J. Analysing collocational patterns of semi-technical words in science textbooks[J]. Pertanika Journal of Social Sciences and Humanities, 2010, 18(2): 241-258.
- [9]SMITH A, DAVIS A. Disambiguating the use of common terms across related medical fields: the problem of intervention[J]. Lexicography, 2018, 4(1): 63-80.
- [10]尉桂英.如何使医学生建立起专业英语词汇体系[J].医学教育, 2002(5): 55-56.
- [11]许家金,吴良平.基于网络的第四代语料库分析工具CQPweb及应用实例[J].外语电化教学, 2014(5): 10-15, 56.
- [12]冯欣,吴菁菁,齐晖,等. MedAca 医学学术英语语料库的创建[J]. 语料库语言学, 2017, 4(2): 106-113, 116.
- [13]王莉莉.基于语料库的农业英语高频动词的语义韵研究[D].咸阳:西北农林科技大学, 2014.
- [14]WEST M. A general service list of English words: with semantic frequencies and a supplementary word-list for the writing of popular science and technology[M]. Lon-

- don; Longmans, Green and Co., 1953.
- [15] COXHEAD A. A new academic word list[J]. TESOL Quarterly, 2000, 34(2): 213-238.
- [16] GARDNER D, DAVISE M. A new academic vocabulary list[J]. Applied Linguistics, 2014, 35(3): 305-327.
- [17] 梁茂成, 李文中, 许家金. 语料库应用教程[M]. 北京: 外语教学与研究出版社, 2010.
- [18] 刘继民. 联想法教学在医学英语词汇教学中的应用[J]. 西北医学教育, 2008, 16(5): 984-985.
- [19] SINCLAIR J, RENOUEF A. A lexical syllabus for language learning[C]. London: Longman, 1988: 140-160.
- [20] 何安平. 语料库语言学与英语教学[M]. 北京: 外语教学与研究出版社, 2004.
- [21] SINCLAIR J. Reading concordances[M]. London: Pearson, 2003.

(编辑: 陈越, 李鑫梅)

[简讯]

校党委常委、纪委书记、监察专员柯瑞清 赴 5 家附属医院开展党风廉政专题教育

近日,按照学校党委部署,校党委常委、纪委书记、监察专员柯瑞清赴附属协和医院、附属第一医院、附属第二医院、附属口腔医院、孟超肝胆医院等 5 家直属(非直属)附属医院作了题为《明大势 守底线 确保白衣天使永保圣洁》的廉政教育专题报告。

柯瑞清结合自身多年从事纪检监察工作的专业实践,围绕卫生健康行业特点,从“党风廉政建设和反腐败斗争的新形势”“《监察法》及监察体制改革对医疗机构的影响”“医疗行业违纪违法案件的特点”“增强‘四个意识’,确保白衣天使永保圣洁”等四个方面,详细解析医疗领域的廉洁政策法规、《监察法》和监察体制改革,并结合医疗领域反面典型案例,为 5 家直属(非直属)附属医院有关人员上了一堂内涵丰富、深入浅出、富有哲理、生动深刻的廉政专题教育课,教育引导每一位同志守好廉洁底线、谨防被“围猎”、牢记“四个没有”、算好人生“七笔账”。

5 家直属(非直属)附属医院共计 880 余人参加了专题讲座,其中医院党政领导班子 40 人,科室主任(副主任)413 人,支部书记 129 人,纪检干部 46 人,其他中层干部 255 人。