

通用词表及其应用研究



罗桂花*, 宋佳苗

中南林业科技大学外国语学院, 湖南长沙 410000

摘要: 通用高频词汇和通用学术词汇是英语核心词汇的重要组成部分, 特别是在 EFL 语境中。通用高频词表提供的是通用核心词汇, 而通用学术词表则提供在广泛的学术领域中出现最核心的词汇。本研究采用文献分析法, 对从 Scopus 数据库以及谷歌学术补检索了通用词表相关文献 5083 条, 文献分析结果表明: 1) 应用较广泛的通用高频词表有 GSL、BNC/COCA2000 和 NGSL, 通用学术词表中 UWL、AWL、AVL 和 NAWL 应用较广。2) 语料库的选择、词汇筛选标准和计词方式是词表开发的关键指标。3) 通用高频词表和通用学术词表在 EAP 教学与研究中被广泛运用, 包括创建新词表、尤其是特定主题学术词表和技术词表, 词汇概貌分析、词汇学习负担分析、教材研发、词汇测试和学习工具开发等领域。4) 词汇覆盖率分析是词表应用的主要方法和途径, 可采用 Compleat Lexical Tutor、RANGE 和 AntWordProfiler 等分析工具。

关键词: 通用高频词表; 通用学术词表; 词表应用; 词汇覆盖率

DOI: [10.57237/j.edu.2023.05.002](https://doi.org/10.57237/j.edu.2023.05.002)

Research on General Word List and Its Application

Luo Guihua*, Song Jiamiao

Foreign Languages College, Central South University of Forestry and Technology, Changsha 410000, China

Abstract: General high-frequency vocabulary and general academic vocabulary are important components of English core vocabulary, especially in EFL context. General high-frequency word list provides the general core vocabulary, while general academic word list provides the most core vocabulary that appears in a wide range of academic fields. By using literature analysis method, this study analyzed 5083 research articles related to the general word list retrieved from Scopus and Google scholar, and the results shows that: 1) GSL, BNC/COCA2000 and NGSL are widely used general high-frequency word lists, while UWL, AWL, AVL and NAWL are widely used in the academic context. 2) Researches on general word lists show that the selection of corpus, the criteria of word and the unit of counting are the key indicators of word list development. 3) General high-frequency word list and general academic word list are widely used in EAP teaching and research, including the creation of new word lists, especially subject-specific academic word list and technical word list, vocabulary profile analysis, vocabulary burden analysis, textbook design, vocabulary testing and vocabulary learning tools development. 4) Coverage analysis is the main method and approach in word list application, which can be conducted by some webs or software such as Compleat Lexical Tutor, RANGE and AntWordProfiler.

Keywords: General High-frequency Word List; General Academic Word List; Word List Application; Lexical Coverage

基金项目: 2022 年湖南省研究生科技创新基金“基于语料库的国内外林学英文学术话语多维度比较研究”(项目编号: CX20220766).

*通信作者: 罗桂花, luowenting8552@126.com

收稿日期: 2023-09-28; 接受日期: 2023-10-25; 在线出版日期: 2023-10-28

<http://www.educationrd.com>

1 引言

学术英语 (EAP) 的教学与研究越来越受到广泛的关注, 而词汇学习和教学是其中最具挑战性的方面之一。EAP 学习者需要哪些词汇? Basturkmen[1]提出英语有一个所有学习者都应该知道的核心核心单词, 而一旦学习者建立了共同核心, 学术化就开始了。而比共同核心单词在非学术领域出现频率低, 但在广泛的学术领域中出现频率较高的学术词汇给学习者带来了更大的困难[2]。学者们通过手工或计算机编制了各种词表, 以识别学术词汇中最有用的词汇。针对共同核心英语的通用高频词表, 以及针对跨学科学术核心词汇的通用学术词表已引起中国学者讨论, 包括: 吴瑾、王同顺[3]讨论了 AWL (Academic Word List) [2] 在上海交大科技英语语料库 (JDEST) 中的适用性。游金干、何家宁[4]对比了三大通用学术词表 UWL (University Word List) [5]、AWL 和 AVL (Academic Vocabulary List) [6] 的选词标准和分级标准, 并探析其不足以建立更为合理的标准。孙凤兰[7]介绍了通用高频词表 GSL, 和通用学术词表 AWL 和 AVL, 并讨论了学术英语口语中学术词表的覆盖率问题。刘迪麟、

雷蕾[8]则专门综述了三个通用学术词表 AWL、AVL 和 ASWL (Academic Spoken Word List) [9]及其相关问题。

已有综述与引介为中国词表研究和应用提供了启示, 也为学术英语教学提供了参考, 但主要聚焦于几个常见通用学术词表综, 通用高频词表仅涉及 GSL, 另外已有研究仅综述了词表基本内容和主要特点, 没有涉及应用问题。鉴于近年词表开发和词表应用研究迅猛发展, 以及学界对通用词表在 EAP 教学与学习中的应用价值的一致认同[10, 11], 本研究在系统介绍通用词表基础上, 试图回答: (1) 哪些通用词表在语言学习和教学中被广泛使用? (2) 通用词表间有什么区别? (3) 通用词表的应用情况如何? 本文分别以主题词“general word list (通用词表)”、“general vocabulary list (通用词表)”、“general academic word list (通用学术词表)”和“general academic vocabulary list (通用学术词表)”于 Scopus 数据库进行检索, 并通过文献链接从谷歌学术补充相关文献, 共获取通用词表相关文献 5083 条。

表 1 通用高频词表

词表	年份	语料库规模	计词方式	词表大小	词汇筛选标准
GSL	1953	500 万	词族	1000*2	词频; 学习容易度; 必要性; 覆盖率; 风格; 情感中立性
BNC/COCA2000	2012	1000 万	词族	1000*2	词频; 广泛分布
NGSL	2014	2.73 亿	修改的词位	2801	词频; 离散度; 标准化词频指数

2 通用高频词表

学术词汇研究中, “共核假设”认为词汇习得的顺序应该是先掌握通用核心词汇, 再学习学术通用词汇[12], 通用高频词表提供的即是通用核心词汇, 文献检索发现常用通用高频词表 7 个¹, 谷歌学术中被引频次较高的有 GSL (2881 次)、new-GSL [15] (361 次)、BNC/COCA2000 [16] (256 次) 和 NGSL [17] (241 次)。进一步文献梳理发现 new-GSL 应用研究非常少, 因此没有纳入介绍。其他三个的概况见表 1。

通用词表中, GSL 和 BNC/COCA 的词汇计数单位都为词族 (word family), 即包括词头的所有屈折

和派生形式; GSL 共有 2000 个词头 (headwords) 及其词族成员[18]。NGSL 的计词方式采用的是修改的词位 (modified lexeme), 即以词头为中心, 包括词头的所有屈折形式, 且不区分词性, 共 2818 个词位²。GSL 词汇筛选标准中有词频 (frequency) 这一客观标准, 和五个主观标准[18]。BNC/COCA2000 沿用了词频这一标准, 并加入了广泛分布 (range)。而 NGSL, 除词频外, 另加了离散度 (Carroll's measure of dispersion) 和标准化词频指数 (standard frequency index, SFI) [19]。

三个词表间最大的差异在于语料库, GSL 的 500 万字书面语料库, 大多数文本在 1930 年之前出版, 被认为是过时的, 以现在标准来看规模也过小[15]。

1 除了文中列出的四个词表, 另还有 BNC [10]、EWL (Essential Word List) [13] 和 NFL7 (Nuclear Family List 7) [14]。

2 Browne[15]去除了其中 19 个词位, 又另增了 2 个词位, 共 2801 个词位。

BNC/COCA2000 是 BNC/COCA 词表³中的前 2000 个词族, 创建于一个专门设计的 1000 万字语料库, 来自于 BNC、COCA 和惠灵顿新西兰英语口语语料库 (the Wellington Corpus of Spoken New Zealand English, WSC), 包括 600 万字英语口语和 400 万字书面英语 [16]。NGSL 的语料库规模最大, 达到 2.73 亿词, 采用的是 16 亿词规模剑桥英语语料库 (The Cambridge International Corpus, CEC) 中的英、美书面和口语文本, 以及剑桥学习者子语料库 [19]。

因为语料库规模和词汇筛选标准的强主观性, GSL 引起了一些争议 [8, 20]。但是, GSL 不仅直接影响了核心英语词汇概念化的方式, 词频、广泛分布和标记覆盖率 (token coverage) 等仍然是创建词表的核心指标, 而且被整合于多个词汇分析工具, 如 Compleat Lexical Tutor [21]、RANGE [22] 和 AntWordProfiler⁴, 广泛用于教学实践和词汇研究。BNC/COCA2000 被认为有更多有用的词汇, 也为学习者所更熟知, 是对第二语言 (L2) 学习者更有用的通用高频词表 [23]。NGSL 在语料库规模和时效性方面较 GSL 进步, 但修改的词位, 较之于词族, 不包括派生形式, 但不区分词性, 导致一个词位含有不同意义, 存在一定的局限性。Browne [17] 在 CEC 语料库中将 NGSL 与 GSL、new-GSL 进行覆盖率测试比较, 证明 NGSL 综合表现更优。

3 通用学术词表

根据学术谷歌数据, 通用学术词表被引频率显示 AWL (4754 次)、AVL (845 次)、UWL (779 次)、AKL [24] (373 次), 以及 NAWL [19] (85 次)。文献阅读发现 AKL 基本上没有应用研究, 因此不纳入分析范围。其他四个概况见表 2。

最早被广泛使用的通用学术词表 UWL 通过合并编辑四个单词列表而创建, 所以缺乏一致的词汇选择原则, 未能设定分布阈值和频数阈值 [4]。在合并四个早期词表后, UWL 排除 GSL 中的 2000 词族, 得到 836

个词族 [5]。

AWL 在 GSL 基础上建成, 共包含 570 个词族, 沿用了词族的计词方式, 以及词频和广泛分布的词汇筛选标准, 词频要求出现频次大于 100 次⁵; 广泛分布则要求在四个子语料库中分别出现 10 次以上, 且需在 28 个子学科领域中的 15 个以上出现; 专业性 (specialised occurrence) 即是排除 GSL 词汇, 以此来区分通用高频词汇和通用学术词汇; AWL 语料库规模为 350 万, 由人文、商业、法律及科学四个学科的 28 个领域的教材、论文、书章节和实验指南构成 [2]。

NAWL, 作为 NGSL 同系列词表, 沿用了修订的词位这一计词方式, 共含 963 个词位。词汇筛选标准仅明确一点, 即通过排除 NGSL 来区分出通用高频词汇, 其他标准不得而知; NAWL 创建于 2.88 亿词的学术语料库, 由剑桥英语语料库 (Cambridge English Corpus)、密歇根学术英语口语语料库 (the Michigan Corpus of Academic Spoken English, MICASE)、英国学术英语口语语料库 (the British Academic Written Corpus, BASE), 以及增补的教材文本组成 [19]。

AVL 是一个独立的学术词表, 不以任何词表为基础, 计词方式为词元 (lemma), 即具有相同词干的所有屈折形式, 且区分词性, 共包含 3015 个词元。AVL 采用 COCA 语料库 (4.25 亿) 中的学术子库, 规模达 1.2 亿词, 共包括九大学科。词汇筛选标准沿用了广泛分布, 具体要求词元至少在九个子语料库中的七个出现, 且出现频率至少达到期望频率的 20%; 另外采用了离散度 (Juilland's D), 要求均匀分布指数需大于等于 0.8; 新增了频次比率 (ratio) 和学科指数 (discipline measure), 前者要求词元在 COCA 学术语料库中出现的频次是其在 COCA 非学术语料库中出现的频次的 1.5 倍, 这样将通用高频词汇和学术高频词汇区分开来; 后者要求词元在九个学科中的出现频率不能超过期望频率的三倍, 以此排除特定学科学术词汇和技术词汇 [6]。

UWL 所合并的四项研究都是在没有计算机辅助下进行的, 语料库较小, 所涉及学科领域也偏少 [2]。AWL 不仅引领了学术词表开发的潮流, 而且被广泛运用于 EAP 课堂教学、语言测试和教材开发, 成为最有影响力的通用学术词表 [25]。AWL 存在两点争议, 一是计词方式为词族, 另一个是 AWL 含有大量通用高频词, 有 451 个词族出现在 BNC 的前 4000 个高频词

3 BNC/COCA 包括 29 个子表, 前 25 个词表各包含 1000 个词族, 另有 4 个补充词表, 即专有名词列表 (proper names)、首字母缩略列表 (acronyms)、复合词列表 (compounds), 和由脏话 (swear words)、感叹词 (exclamations) 和字母词 (letters of the alphabet) 组成的边缘词列表 (marginal words)。

4 详见 <https://www.laurenceanthony.net/software>

5 有一个词族除外, 仅 80 次, 因为该词族仅一个单词。

族中, 这是因为 AWL 简单地以排除 GSL 来区分通用高频词和学术高频词而导致[6]。Gardner & Davies [6] 将 AVL 与 AWL 在 COCA 和 BNC 的学术语料库中进行比较证明 AVL 覆盖率高于 AWL。虽然 AVL 的新方

法逐渐替代了 AWL 成为学术词表开发的参考标准[26], 但其因语料库单一, 仅采用美国英语语料, 且未收录教科书和学术口语语料, 另外子语料库之间的不平衡也受到质疑[27]。

表 2 通用学术词表

词表	年份	语料库规模	计词方式	词表大小	词汇筛选标准
UWL	1984	未知	词族	836	不统一
AWL	2000	350 万	词族	570	词频; 广泛分布; 专业性
AVL	2014	1.2 亿	词元	3000	广泛分布; 离散度; 频次比率; 学科指数
NAWL	2013	2.88 亿	修改的词位	963	未知

4 词表的应用

近年来, 词表应用也得到越来越多的关注, 尤其是在 EAP 教学与学习语境中。已有通用词表不仅促成了新词表, 尤其是特定学科学术词汇 (discipline-specific academic words) 和技术词汇

(technical words) 不断涌现, 在确立 EAP 词汇学习目标、评估词汇知识和增长、分析文本难度和丰富度、创建和修订 EAP 教学材料、设计词汇学习工具、开发 EAP 课程等方面都发挥效用[6]。通过对文献进行系统回顾与总结, 通用词表在 EAP 词汇学习与教学中的应用可概括为图 1 所示。

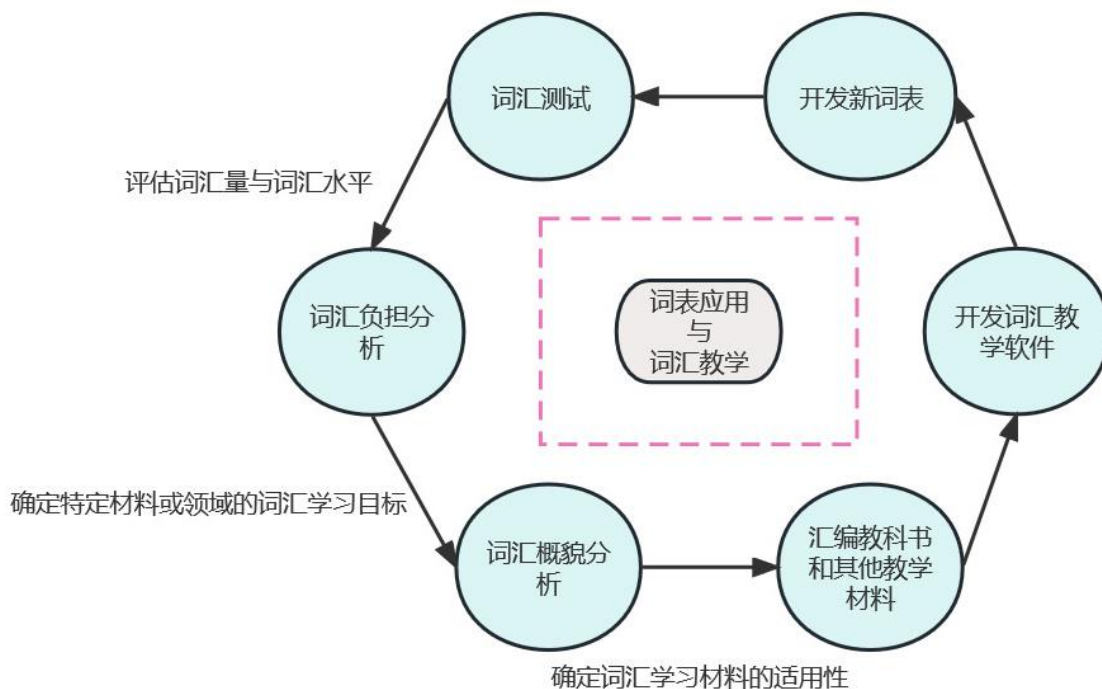


图 1 通用词表的应用

4.1 作为新词表创建的基准

创建新词表的一种选择是以已有词表为基础, 并再分为两种情况, 一种情况是 EWL 和 NFL7 以新的词汇筛选标准来精选或删减已有词表而形成, EWL 是在 GSL、BNC2000、BNC/COCA2000 和 new-GSL 四大词表的基础上创建[26], NFL7 则将 BNC/COCA 词表

的前 3000 词族的 19062 词形(word types)减少到 7293 而形成[14]。另一种情况是, AWL 和 NAWL 分别在 GSL 和 NGSL 基础上创建, 即 GSL 和 NGSL 被用作区分通用高频词汇和学术词汇的基线而被排除在 AWL 和 NAWL 之外。这种选择在早期特定学科学术词表和技术词表的创建中也被沿用, 如 Wang, Liang & Ge [28]的医学学术词表排除了词表中与 GSL 重叠的

词汇; Mukundan & Jin [29]的护理学技术词表以 AWL、GSL 为基础。

创建新词表的另一种选择不以已有词表为基础, 从头开始创建, 比如 AVL, 它不是简单地排除通用高频词表来区分通用高频词和学术高频词, 而是通过频次比率等标准来区分。这种选择的另一典范是 Lei & Liu [26]开发的医学学术词表 (MAVL), 因为不少通用高频词在学术英语中的频率比通用英语高得多, 且在学术英语里通常具有特殊意义, 所以, MAVL 使用已有通用高频词表 (new-GSL) 来识别通用高频词, 再确定高频词是否具有医学意义, 然后结合其他标准来区分出通用高频词。如何识别通用高频词是否具有学术意义或者是特定学科意义是特定学科学术词表的一个关键问题, 简单地以 GSL、AWL 等通用词表为基准来识别通用高频词和通用学术词从而加以排除的方法已经被质疑和淘汰。但这个问题的解决还是有赖于通用词表, 比如 MAVL 利用了 new-GSL, 又比如 Roesler [30]在创建计算机科学学术词表 (CSAVL) 时使用 NGSL, 结合牛津计算机科学词典来识别具有特定学术意义的通用高频词。

另外, 在开发新词表时, 已有通用词表常用于比较以显示新词表的有效性和特征。其中, 词表覆盖率 (lexical coverage) 测试是验证词表有效性的通用方法 [13], 也就是说, 已有词表被用作比较对象和标准, 如前文所述, 为了证明 NAWL 和 AVL 有效性, Browne [17] 和 Gardner & Davies [6] 分别展开了覆盖率比较。另有 Tongpoon-Patanasorn [31]将新创建的金融技术词表与 GSL、AWL 进行比较, 显示 413 个词形 (42.18%) 与 GSL 重合, 291 个词形 (29.72%) 与 AWL 重合。

可以说, 通用词表在创建新词表的过程中具有多重作用, 可作为新词表精选的基础, 可用于区分通用高频词汇和学术词汇, 又用作比较对象和标准显示新词表的有效性和特征; 不仅可用于创建新的通用高频词表和通用学术词表, 在特定学科学术词表和技术词表的开发中也是重要工具和参考。

4.2 词汇概貌分析

词汇覆盖率指特定词表在语料库中覆盖的词汇百分比, 可用文本分析工具 Compleat Lexical Tutor、RANGE 或 AntWordProfiler 等计算得出 [32]。除了用于上文所提到的验证词表有效性, 词汇覆盖率还可用于词汇概貌分析, 即调查和描述不同 ESP 或者 EAP 文本

类型的词汇状况, 提供关于文本词汇的难易度和代表性, 以及文本词汇在词表不同频率层级的分布等数据 [32]。Chen & Ge [33]结合体裁语步分析显示 AWL 在医学研究论文及其五个语步中的覆盖率均为 10% 左右, 292 个 (51.2%) AWL 词族被广泛使用。Li & Qian [34]利用 RANGE 分析了 630 万词的香港金融服务语料库 (HKFSC) 的词汇概貌, 发现 AWL 在不同文本类型中覆盖率有明显变化 (8.01%-19.73%), 平均覆盖率为 10.46%。Jahangard et al. [35]分析了 AWL 和 GSL 在 400 万词硬科学语料库中的覆盖率, 结果显示 GSL 覆盖率约为 78.4%, AWL 约为 16%, 证明所构建的硬科学语料库的有效性。Liu & Chen [36]利用 AntWordProfiler、ASWL 和 BNC/COCA2000 分析包含六大主题、约 437 万字的 TED 演讲语料库的词汇概貌, 结果显示, BNC/COCA2000 与 ASWL 的 1741 个词族的覆盖率相近, 表明 ASWL 更适合 TED 演讲学习, TED 适合作为学术听力教学材料。Skoufaki & Petrić [37]词汇概貌分析显示有近三分之一的 AVL 词汇出现在英国 EAP 课程学习材料中, 且其中大多数是 AVL 的前 1000 个词元。

4.3 词汇负担分析

词汇覆盖率还常用于调查理解不同类型文本所需的词汇量, 为教师和学习者提供理解特定文本所需的目标词汇量, 即词汇负担分析 (vocabulary load) [10]。这方面的研究建立在实证研究基础上, 最有影响的是 Laufer & Ravenhorst-Kalovski [38]所提出的: 词汇覆盖率达到 95% 是阅读理解阈值最低标准, 98% 覆盖率是独立阅读的充分理解阈值。95% 和 98% 理解阈值被广泛用于计算词汇负担, 即第二语言 (L2) 学习者需要多少词汇量达到基本或充分理解某一类或某一领域文本。一方面, ESP 教材和学术论文词汇量负担分析广受关注。Hsu [39]利用 RANGE 和 BNC/COCA 分析了工程类英语教科书的词汇负担, 前 5000 个词族加上四个补充词表可达到工程类教科书 95% 的覆盖率。Nguy & Ha [40]对 10 个主要科学学科约 10 万篇学术摘要共 2600 万单词的词汇负担分析表明要达到 95% 和 98% 的文本覆盖率分别需要 BNC/COCA 中前 7000 和前 15000 个词族分别加上 4 个补充词表。另一方面, 其他常见 EAP 学习材料和学习场景的词汇负担分析也有所涉及。Nurmukhamedov & Sharakhimov [41]对英语播客的

研究显示 BNC/COCA 前 3000 个和前 5000 个词族分别加上 4 个补充词表可以达到 96.75% 和 98.26% 的覆盖率。Xodabande, Ebrahimi & Karimpour [42] 利用 AntWordProfiler 分析了约 445 万规模的慕课(MOOC) 语料, 结果显示, BNC/COCA 词表前 5000 和前 9000 词族可以分别达到 95% 和 98% 的覆盖率, 而 GSL 仅达到 85.54%, AWL 仅 5.32%。Dang [43] 利用 RANGE 软件分析了 20 个学术领域 104 场会议报告的近 57 万字语料, 结果显示 BNC/COCA 词表前 3000 和前 5000 个词族分别加上专有名词表、边缘词表可以达到 96.84% 和 98.22% 的覆盖率。

4.4 研发教材及教辅材料

为了教师和学习者选择、确定合适的 EAP 学习材料和资源, 一些出版物将重点放在一个或多个特定词表上, 尤其是教材和字典。教材中, 剑桥大学出版社新出版的一套三级教科书《聚焦》(*In focus*) 着重于 NGSL 和 NAWL 的系统教学。密歇根大学出版社的《掌握: UWL 词表读本》(*Mastery: A University Word List Reader*) 基于 UWL 突出美国学术语境中最常用的词汇。朗文出版社《词汇力量》(*Vocabulary power*) 系列专注最重要的单词, 三本书每本 300 个新单词, 即 GSL 的 500 个单词和 AWL 的 400 个单词。有大量的教材专注于 AWL, 包括朗文出版社的《词汇聚焦: 掌握 AWL 词表》(*Focus on Vocabulary: Mastering the Academic Word List*)、牛津大学出版社的五年级阅读教材《学术阅读系列》(*Inside reading*) 系列、剑桥大学出版社的《使用中的学术词汇》(*Academic Vocabulary in Use*)、霍顿米夫林出版公司出版的《基本学术词汇: 掌握完整的学术词汇表》(*Essential Academic Vocabulary: Mastering the Complete Academic Word List*) 和《学术英语成就系列》(*English for Academic Success Series*) 等。词典中, 朗文考试词典、朗文当代英语词典、牛津高级学习词典和牛津学生英语词典等中也突出强调了 AWL 词汇[44]。包括和强调学术词汇已经成为教科书和词典的一个重要发展领域, 以帮助 EAP 学习者的学习[45]。

4.5 开发词汇测试工具

如何对学习者的词汇量进行科学有效的测试一直是一个热门研究问题, 尤其是 EFL 语境中的 EAP 学习者, 词汇水平参差不齐。随着词表的开发, 各种基于词表

的词汇量测试工具被开发以确定 EAP 学习者的出发点。二语书面接受性词汇知识测试使用最广泛的工具——词汇水平测试 (VLT) [46] 以非连续性词频表大体估测学习者对常见英语词汇含义的掌握, 包含 5 个测量工具测试 2000、3000、5000、10000 词频以及常见学术词汇的了解, 其中 2000 词频和学术词汇就是基于 GSL 和 UWL 开发的。后续 Schmitt, Schmitt & Clapham [47] 版本的 VLT 中学术词汇知识的测试由 UWL 改为了 AWL; McLean & Kramer [48] 和 Webb, Sasao & Balance [49] 的版本都采用了 BNC/COCA 词表中 1000-5000 词族, 前者还含有 AWL 学术词汇知识测试。

Stoeckel & Bennett [50] 基于 NGSL 研制的 NGSLT (New General Service List Test) 和 Stoeckel & Bennett [51] 基于 NAWL 研制的 NAWLT (New Academic Word List Test), 以及基于 AWL 开发的 AVST (the Academic Vocabulary Size Test) [52] 和 AVT (the Academic Vocabulary Test) [53] 都是接受性词汇测试工具。另外, McLean, Kramer & Beglar [54] 基于 BNC/COCA 和 AWL 还创建了听力词汇水平测试 (Listening Vocabulary Levels Test, LVLTL), 采用 VST 范式来测量 BNC/COCA 前 5000 词族和 AWL 的词汇水平知识。

基于通用词表创建的产出性词汇量测试中较有代表性的是 LFP (the Lexical Frequency Profile) [55], 利用 GSL1000、GSL2000、UWL 三个通用词表来测量学习者使用不同层次词汇的水平。深度词汇测试中, 有 Read [56] 基于 UWL 创建的词汇联想测试 WAT (Word Association Test); Read & Dang [57] 基于 AVL 开发的新词汇深度测量方法, 包括对同义词、搭配和词类的接受性知识的测量。

4.6 开发词汇教学工具

词表的快速发展也催生了各种词汇自主学习和词汇教学工具和软件的开发和使用。首先是词会移动学习软件, 比如基于“间隔重复”的科学原理, NGSL Builder、NAWL Builder 和 AWL Builder 都包含智能词表制作器、智能闪卡系统, 以及英语母语专业语音人士对每个单词的发音等工具可以帮助学习者快速有效地学习 NGSL、NAWL 或者 AWL 单词, 并可以下载、发送电子邮件和组织 NGSL、NAWL 或者 AWL [58]。实验表明 NGSL builder 和 NAWL builder 对高频词的短期和长期学习有很大帮助[59, 60], AWL builder 对 EFL 语境中 EAP 词汇自主学习有效性较高[61]。

其次,一系列基于网络的线上工具可供教师展开 EAP 词表教学与分析,比如 AWL Highlighter 可高亮显示文本中的 AWL 词汇,开发填空题以帮助学生练习文本中的 AWL 词汇,或是制成词汇云(tag cloud); AWL Gapmaker 也可以将电子文本中所包含的 AWL 词汇替换为空白供学生练习,有利于写作词汇教学[62]。这两个工具都可以在 EAP 学习网站 eapfoundation⁶上找到,类似的还有 NAWL Highlighter 和 AVL Highlighter 也可用于高亮显示 NAWL 和 AVL 词汇并开发填空题; Vocabulary profiler 则可供我们任选前文提到的大多数通用词表。COCA 提供的互动工具以多种方式促进学术词汇的学习,在线界面⁷帮助分析文本并突出显示 AVL 中的所有学术词汇;还为文本中点击的任何词汇提供详细信息,包括定义、频率和搭配。

另外,由 Tom Cobb 开发的 Compleat Lexical Tutor⁸ 是一个集多个大型语料库的多功能的数据驱动语言学习免费网络平台,被广泛用于 EAP 研究和教学,比如用于前述的词汇概貌(vocabulary profile)和广泛分布(RANGE)分析;还有测试(test)功能,以及词族/词元化(fam/lemizer)功能,即将 GSL/COCA 等以词族为计词方式的词表转化为词元形式等[63]。另外,可用于创建关于 GSL、AWL 等词表的完形填空(VP Cloze Builder)、制作 EAP 词汇闪卡(flashcard),以及生成超文本(Hypertext)等。另还有选词填空(I-D Word identification quiz)可用于对 GSL、AWL 和 BNC/COCA 等词表中的词汇进行辨析训练。

5 结语

通用词表为 EAP 教学和学习,尤其是 EFL 语境中的初级和中级学习者提供了通用核心词汇和通用学术词汇。教师和学习者可以根据需要从几大广泛被运用的通用高频词表到通用学术词表进行选择,词表之间的区别和关键指标主要体现在以下三个方面:(1)语料库选择与建设。除了早期的 GSL 和 UWL,其它词表都基于语料库,且已达到上亿规模(NGSL 2.73 亿,NAWL 2.88 亿规模);GSL、UWL、AWL 和 AVL 仅使用书面语料,BNC/COCA2000、NGSL 和 NAWL 都囊括了口语语料库,AWL 则自建了包括多个学科的语

料库,引发了特定学科学术词表和技术词表的创建。

(2)词汇筛选标准。词频是最一致的词汇筛选标准⁹。广泛分布和离散度很常用,以确保词表中词汇在子语料库中的广泛分布和均匀分布。区分学术词汇和非学术词汇是学术词表筛选词汇的重要步骤,UWL、AWL 依据专业性标准简单排除 GSL,AVL 采用的是更复杂的频次比率和学科指数标准。(3)计词方式。通用词表共使用了三种计词方式,GSL、BNC/COCA2000、UWL 和 AVL 为词族;NGSL 与 NAWL 是修改的词位,AVL 是词元。后二者的区别是修改的词位包括词干及其屈折变化的所有形式,而词元则首先进行词行还原,仅以词干的形式呈现。

已有研究表明,通用高频词表和通用学术词表已被广泛应用于 EAP 词汇教学的整个过程。它们可用于词汇水平测试以评估 EAP 学习者的词汇知识和能力,而词汇负担分析则可以确定特定 EAP 学习材料或领域的词汇目标,词汇概貌分析可以确定学习材料的适应性;新的词表,特别是特定主题学术词表和技术词表的开发也有赖于已有的通用词表,EAP 词汇教学软件、教材和词典等教学资源也不断开发。通用词表应用形成了 EAP 词汇学习的闭环,可以确保学生从所学词汇中获得最大的收益。近年来,一大批英语特定主题学术词表和技术词表被开发,还出现了多词词表、搭配词表等热门方向,但从中国已有研究来看,中国英语词表的开发和应用研究都还非常稀少和有限,大有可为。

参考文献

- [1] BASTURKMEN H, 2006. Ideas and Options in English for Specific Purposes [M]. New Jersey: The University of Auckland.
- [2] COXHEAD A, 2000. A new academic word list [J]. TESOL quarterly 34(2): 213-238.
- [3] 吴瑾,王同顺,2007. Coxhead“学术词汇表”的适用性研究[J]. 国外外语教学, (2): 28-33.
- [4] 游金干,何家宁,2016. 通用学术词表的对比研究——选词、评估和分级标准 [J]. 山东外语教学, (6): 50-58.
- [5] XUE G Y, NATION I S P, 1984. A university word list [J]. Language learning and communication 3(2): 215-229.

6 <https://www.eapfoundation.com/>

7 <https://www.academicvocabulary.info/>

8 <https://www.lex tutor.ca/>

9 除了 AVL。

- [6] GARDNER D, DAVIES M, 2014. A new academic vocabulary list [J]. *Applied linguistics* 35(3): 305-327.
- [7] 孙凤兰, 2017. 国外学术英语口语词汇覆盖率最新研究述评 [J]. *外语研究*, (5): 61-65.
- [8] 刘迪麟, 雷蕾, 2020. 学术词表研究综述 [J]. *外语教学*, (2): 34-38+50.
- [9] DANG T N Y, COXHEAD A, WEBB S, 2017. The academic spoken word list [J]. *Language Learning* 67(4): 959-997.
- [10] NATION I S P, 2006. How large a vocabulary is needed for reading and listening? [J]. *Canadian modern language review* 63(1): 59-82.
- [11] SCHMITT N, COBB T, HORST M, SCHMITT D, 2017. Schmitt. How much vocabulary is needed to use English? Replication of van Zeeland & Schmitt (2012), Nation (2006) and Cobb (2007). *Language Teaching*, 50(2), 212-226.
- [12] BLOOR M, BLOOR T, 1986. Languages for Specific Purposes: Practice and Theory [J]. *Occasional Paper-Centre for Language and Communication Studies*, (19): 1-35.
- [13] DANG T N Y, WEBB S, 2016. Making an essential word list for beginners. In P. Nation (Ed.), *Making and Using Word Lists for Language Learning and Testing*. Amsterdam: John Benjamins.
- [14] COBB T, LAUFER B, 2021. The nuclear word family list: A list of the most frequent family members, including base and affixed words [J]. *Language Learning* 71(3): 834-871.
- [15] BREZINA V, GABLASOVA D, 2015. Is there a core general vocabulary? Introducing the new general service list [J]. *Applied Linguistics*, 36(1): 1-22.
- [16] NATION I S P, 2012. The BNC/COCA Word Family Lists [EB/OL]. Available online: <http://www.victoria.ac.nz/lals/about/staff/paulnation>.
- [17] BROWNE C, 2014. A new general service list: The better mousetrap we've been looking for [J]. *Vocabulary learning and Instruction*, 3(2): 1-10.
- [18] WEST M, 1953. A general service list of English words: with semantic frequencies and a supplementary word-list for the writing of popular science and technology [M]. Longman.
- [19] BROWNE C, Culligan B, Phillips J, 2013. The new academic word list [EB/OL]. Retrieved from www.newacademicwordlist.org.
- [20] SCHMITT N, MEARA P, 1997. Researching vocabulary through a word knowledge framework: Word associations and verbal suffixes. *Studies in second language acquisition* 19(1): 17-36.
- [21] COBB T, FREE P, 2004. The Compleat lexical tutor, v. 4. TESL-EJ 8(3).
- [22] HEATLEY A, NATION P, COXHEAD A, 2004. RANGE. Wellington: Victoria University of Wellington. Available from <http://www.victoria.ac.nz/lals/staff/paul-nation.aspx>.
- [23] DANG T N Y, WEBB S, COXHEAD A, 2022. Evaluating lists of high-frequency words: Teachers' and learners' perspectives [J]. *Language Teaching Research* 26(4): 617-641.
- [24] PAQUOT M, 2010. Academic vocabulary in learner writing: From extraction to analysis [M]. Bloomsbury Publishing.
- [25] 马蓉, 2017. 学术词汇研究四十五年 [J]. *现代外语*, (3): 420-428+439.
- [26] LEI L, LIU D L, 2016. A new medical academic word list: A corpus-based study with enhanced methodology [J]. *Journal of English for academic purposes* 22: 42-53.
- [27] GHOLAMINEJAD R, SARAB M R A, 2021. A comparison of the Academic Word List and the Academic Vocabulary List: Should the AVL replace the AWL? [J]. *TEFLIN Journal* 32(1): 167-182.
- [28] WANG J, LIANG S L, GE G C, 2008. Establishment of a medical academic word list [J]. *English for Specific Purposes* 27(4): 442-458.
- [29] MUKUNDAN J, JIN N Y, 2012. Development of a technical nursing education word list (NEWL) [J]. *International Journal of Innovation in English Language Teaching and Research* 1(2): 105.
- [30] ROESLER D, 2021. When a bug is not a bug: An introduction to the computer science academic vocabulary list [J]. *Journal of English for Academic Purposes* 54: 101044.
- [31] TONGPOON-PATANASORN A, 2018. Developing a frequent technical words list for finance: A hybrid approach [J]. *English for Specific Purposes* 51: 45-54.
- [32] PARIBAKHT T S, WEBB S, 2016. The relationship between academic vocabulary coverage and scores on a standardized English proficiency test [J]. *Journal of English for Academic Purposes* 21: 121-132.
- [33] CHEN Q, GE G C, 2007. A corpus-based lexical study on frequency and distribution of Coxhead's AWL word families in medical research articles (RAs) [J]. *English for Specific Purposes* 26(4): 502-514.
- [34] LI Y Y, QIAN D D, 2010. Profiling the Academic Word List (AWL) in a financial corpus [J]. *System* 38(3): 402-411.

- [35] JAHANGARD A, RAJABI A, KHALAJI K, CHOLMAGHANI A, 2015. An investigation of academic word list (AWL) and general service list (GSL) in hard sciences' research articles: Meeting the rising demands in ESP [J]. *Journal of Teaching English for Specific and Academic Purposes* 2(4): 603-614.
- [36] LIU C Y, CHEN H H J, 2019. Academic spoken vocabulary in TED Talks: Implications for academic listening [J]. *English Teaching & Learning* 43: 353-368.
- [37] SKOUFAKI S, PETRIĆ B, 2021. Academic vocabulary in an EAP course: Opportunities for incidental learning from printed teaching materials developed in-house [J]. *English for Specific Purposes* 63: 71-85.
- [38] LAUFER B, RAVENHORST-KALOVSKI G C, 2010. Lexical threshold revisited: Lexical text coverage, learners' vocabulary size and reading comprehension [J]. *Reading in a Foreign Language* 22: 15-30.
- [39] HSU W, 2014. Measuring the vocabulary load of engineering textbooks for EFL undergraduates [J]. *English for Specific Purposes* 33: 54-65.
- [40] NGUY N L Q, HA H T, 2022. Lexical Profile of Academic Written English Revisited: What Does it Take to Understand Scholarly Abstracts? [J]. *SAGE Open*, 12(3): 21582440221126342.
- [41] NURMUKHAMEDOV, U, SHARAKHIMOV S, 2021. Corpus-based vocabulary analysis of English podcasts [J]. *RELC Journal* 0033688220979315.
- [42] XODABANDE I, EBRAHIMI H, KARIMPOUR S, 2022. How much vocabulary is needed for comprehension of video lectures in MOOCs: A corpus-based study [J]. *Frontiers in Psychology* 13.
- [43] DANG T N Y, 2022. A corpus-based study of vocabulary in conference presentations [J]. *Journal of English for Academic Purposes* 59: 101144.
- [44] COXHEAD A, 2011. The academic word list 10 years on: Research and teaching implications [J]. *Tesol Quarterly* 45(2): 355-362.
- [45] COXHEAD A, NGUYEN T M H, 2023. Taking a practical approach to academic vocabulary in second language teaching and learning//*Handbook of Practical Second Language Teaching and Learning*. Routledge 421-434.
- [46] NATION J L, 1983. A new method using hexamethyldisilazane for preparation of soft insect tissues for scanning electron microscopy [J]. *Stain technology* 58(6): 347-351.
- [47] SCHMITT N, SCHMITT D, CLAPHAM C, 2001. Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test [J]. *Language testing* 18(1): 55-88.
- [48] MCLEAN S, KRAMER B, 2015. The creation of a new vocabulary levels test [J]. *Shiken* 19(2): 1-11.
- [49] WEBB S, SASAO Y, BALANCE O, 2017. The updated vocabulary levels test: Developing and validating two new forms of the VLT[J]. *ITL—International Journal of Applied Linguistics* 168 (1): 33–69.
- [50] STOECKEL T, BENNETT P, 2015. A test of the new General Service List [J]. *Vocabulary Learning and Instruction* 4(1): 1-8.
- [51] STOECKEL T, BENNETT P, 2020. The New Academic Word List Test (NAWLTL). Available at: <http://www.newgeneralservicelist.org/ngsl-levels-test>.
- [52] MASRAI A, MILTON J, 2018. Measuring the contribution of academic and general vocabulary knowledge to learners' academic achievement [J]. *Journal of English for Academic Purposes* 31: 44-57.
- [53] PECORARI D, SHAW P, MALMSTRÖM H, 2019. Developing a new academic vocabulary test [J]. *Journal of English for Academic Purposes* 39: 59-71.
- [54] MCLEAN S, KRAMER B, BEGLAR D, 2015. The creation and validation of a listening vocabulary levels test [J]. *Language Teaching Research* 19(6): 741-760.
- [55] LAUFER B, NATION P, 1995. Vocabulary size and use: Lexical richness in L2 written production [J]. *Applied linguistics* 16(3): 307-322.
- [56] READ J, 1993. The development of a new measure of L2 vocabulary knowledge [J]. *Language testing* 10(3): 355-371.
- [57] READ J, DANG T N Y, 2022. Measuring depth of academic vocabulary knowledge [J]. *Language Teaching Research* 13621688221105913.
- [58] BROWNE C, 2021. The NGSL Project: Building Wordlists and Resources to help EFL Learners (and Teachers) to Succeed [J]. JALTCALL Publications.
- [59] ZAKIAN M, XODABANDE I, VALIZADEH M, YOUSEFVAND M, 2022. Out-of-the-classroom learning of English vocabulary by EFL learners: investigating the effectiveness of mobile assisted learning with digital flashcards [J]. *Asian-Pacific Journal of Second and Foreign Language Education* 7(1): 1-16.
- [60] XODABANDE I, IRAVI Y, MANSOURI B, MATINPARSA H, 2022. Teaching Academic Words With Digital Flashcards: Investigating the Effectiveness of Mobile-Assisted Vocabulary Learning for University Students [J]. *Frontiers in Psychology* 2903.

- [61] XODABANDE I, ATAI M R, 2022. Using mobile applications for self-directed learning of academic vocabulary among university students [J]. Open Learning: The Journal of Open, Distance and e-Learning 37(4): 330-347.
- [62] COXHEAD A, BYRD P, 2007. Preparing writing teachers to teach the vocabulary and grammar of academic prose [J]. Journal of second language writing 16(3): 129-147.
- [63] 刘玉山, 胡志军, 2012. 基于在线免费语料库 The Compleat Lexical Tutor 的词汇自主学习与评估 [J]. 山东外语教学, (6): 64-69.

作者简介

罗桂花

1982 年生, 副教授, 博士. 研究方向为学术话语分析、法庭话语分析、语料库语言学、ESP 研究.

E-mail: luowenting8552@126.com

宋佳苗

1997 年生, 硕士研究生. 研究方向为语料库语言学、ESP 研究.

E-mail: a594814324@163.com