

绪 论

随着我国高等教育国际化的持续深入，英语作为学术交流与教学媒介语言的趋势不断增强，特别是在“双一流”高校建设背景下，学术英语研究的规模和需求日益增长。根据中国知网检索统计显示，2000年至2017年国内有关学术英语的文献多达2,600篇，特别是2015年、2016年、2017年三年，论文共计约1,350篇，超过这18年的半数。学术英语研究探索学术英语的语言本体特征，包括词汇语法、篇章结构等，并指导学术英语教学的设计和实验，对培养国际化人才具有十分重要的作用。因此，学术英语一直是我国外语教育的热议话题。

在上述2,600篇文献中，约有650篇运用语料库方法进行实证研究，占总数的25%，而且比例不断上升(见图1)。在国外文献中，与语料库相关的学术英语研究也呈日益增长的趋势。谷歌学术检索结果显示(见图2)，在2000年约有550篇文献运用语料库方法考察学术英语的语言特征，2017年数量增长至近五倍。类似地，Web of Science数据库中与语料库相关的学术英语研究论文数量从2000年的20篇增长至2017年的60篇，增长至三倍(见图3)。可见，语料库研究方法已经成为学术英语研究的重要途径。

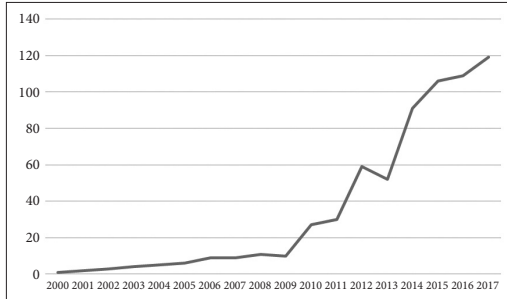


图 1 2000 年至 2017 年中国知网中运用语料库方法的学术英语研究论文数量

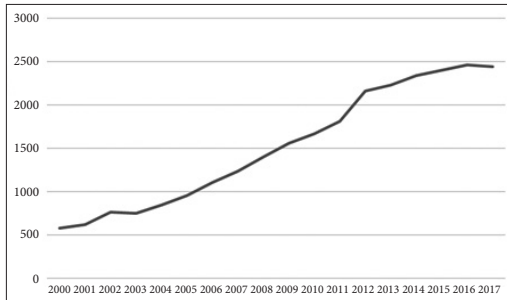


图 2 2000 年至 2017 年谷歌学术中运用语料库方法的学术英语研究文献数量

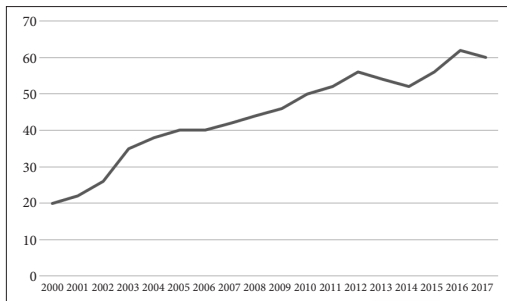


图 3 2000 年至 2017 年 Web of Science 中运用语料库方法的学术英语研究论文数量

值得一提的是,在国内650篇运用语料库方法的学术英语研究文献中,有期刊论文485篇,占总数的74.6%;硕博论文160篇,占总数的24.6%;会议论文5篇,占总数的0.8%。学术英语研究硕博学位论文的比例(特别是硕士论文,占比达19%)明显高于应用语言学其他方向的论文。因此,常常有观点认为,学术英语语料库研究的入门门槛低,是能够轻松开展研究的领域。然而,实际并非如此。当前普遍的问题是,现有研究同质化现象严重,研究者常常机械地套用国外理论或分析框架,未能深入思考学术英语研究的本质、语料库方法的适用性与解释力以及研究结果的教学指导意义。比如,有人认为Hyland(2005c)的立场框架是一个热点,便随意选取某个学术体裁文本建立语料库,分析其立场标记的频率和分布。但是,这种研究实践可能无法回答一些潜在的重要问题,如“为什么要在该体裁中考察立场修辞策略而非其他修辞策略?”“为什么基于Hyland(2005c)的立场概念和框架研究作者的态度,而非基于Hunston(2011)的‘评价’(evaluation)或Martin & White(2005)的‘评估’(appraisal)?”。因此,准确掌握语料库和学术英语的概念与特征,系统梳理学术英语语料库的相关研究,是把握研究创新性、实践性和价值性的前提。

《绪论》中的文献趋势表明，学术英语语料库研究在应用语言学领域具有举足轻重的地位，能够满足新时期英语教学和语言应用的需求，剖析学术交流语境中的英语语言特征，并为学术英语教学提供实证依据。厘清语料库和学术英语的概念与特征，无疑是充分理解学术英语语料库研究理据和目标的重要前提。本章将阐述语料库与学术英语的概念、范式及特征，并探讨二者结合的重要性与必要性。

1.1 语料库的概念与特征

Sinclair (1991: 171) 较早给出具有代表性的语料库定义：“a collection of naturally occurring language text, chosen to characterize a state or variety of a language” [(语料库是) 自然发生的语言文本汇集，用于描写某种语言的状态和变体]。Hunston (2002: 2) 更为具体地将语料库界定为“a collection of naturally occurring examples of language, consisting of anything from a few sentences to a set of written texts or tape recordings, which have been collected for linguistic study” [(语料库是) 为语言学研究而收集的自然发生的语言集合，可以是数个句子，也可以是文本或录音

集]。由上述定义可见,语料库为语言描写提供了丰富的语言实例,从而更好地帮助我们观察自然语言使用的规律和特点。

杨惠中(2002)总结道,语言学研究必然涉及语言材料,根据采集和使用语言材料的不同途径,现代语言学研究的基本方法主要有三种,即内省法(introspection)、诱导法(elicitation)和语料库相关的方法(corpus-related approach)。内省法以语言学家本人为语料提供者,以依靠语言学家自己的语感作为判断歧义、正误、可接受性等的依据。诱导法是通过实地调查收集人们对实际使用的语言材料的看法和心理反应,通常采用有控制的方法诱导出被试者对句子或句子中某个成分的判断,要求被试者确定句子中是否有错误、句子的可接受程度、其对句子的理解程度以及其他类似的有关数据。采用诱导法可以使结论带有某种程度的客观性和可靠性,从而获得对某个语言事实可接受程度的判断。语料库方法则是在随机采样的基础上,收集具有代表性的真实语言材料。可以说,研究语料库中的语言材料近似于研究语言本身。

语料库样本多、规模大、来源广的特点使得语言研究及其结论具有很强的客观性、普遍性与稳定性,所得到的结论不会被轻易否定。因为“比较大的样本可以减少样本统计量的变异”(崔希亮、张宝林 2011: 103),所以语料库的优越性能够保证结论的可靠性。正如梁茂成(2016: 6-7)的评价:从某种意义上说,语料库是语言理论的试金石。如果我们提出某种理论或者利用某种方法得出一种结论,但这种结论与语料库呈现的语言事实不符,我们则有理由认为,这种理论还需要进一步完善。

但是,语料库的不足之处是无法告诉我们文本作者使用某一特定语言形式的原因。如Cook(1998)所言,语料库不能告知研究者语言如何在大脑中得到处理或者无法揭示某一语言结构的认知选择。这也意味着,研究者需要对语料库的发现补充文本参与者或代表群体的认知和识解;另一方面,无论多大规模的语料库都无法囊括语言使用的全部。“语料库不过是片面实据(Corpora are only partial authorities.)。语言使用者的语言经

验随着积累而日益增长和丰富”(Cook 1998: 39)。因此,我们需要认识到,语料库是在一定时期内收集的数据片段,任何语料库都具有局部时间性,在建立之后可能很快就会过时。但是这并不意味着基于语料库的发现是无效的,只不过在对发现下结论时需要谨慎,把握好结论适用的度和范围。此外还应该认识到,对于语言形式和表现缺位的情况,语料库常常显得无能为力。换句话说,语料库能够告诉我们语料文本中有什么语言,却难以告知不含有什么语言。在应用语料库发现时,我们应该充分认识到这一局限。例如,基于多学科英语学术论文语料库的研究发现自然和生命科学领域论文使用立场标记的频率明显少于人文和社会科学领域,但是我们不能草率地认定理工学科学术论文缺少学术立场和判断,只不过理工科语篇很可能不是通过显性的语言形式建构立场,而是基于隐性知识表达价值判断。

1.2 语料库研究的主要范式

语料库研究的主要范式根据对语料库语言学的定位和视角不同大体分为“基于语料库的范式”(corpus-based approach)和“语料库驱动的范式”(corpus-driven approach)。前者突出语料库语言学的工具性,认为语料库为语言研究提供大量可靠的素材和实据;后者强调语料库语言学的理论性,认为它不仅是工具和方法,更是独立于其他已有语言学理论,从词语和短语学视角描写和分析语言的一种理论(梁茂成 2016)。

1.2.1 基于语料库研究范式

在基于语料库(corpus-based)的研究范式中,语料库被用作数据库,目的是把语料库作为事实依据,验证、示例或解释已有理论,这种范式不是从语料库中发现或概括出新的理论(梁茂成 2016)。较早作出两类语料库研究范式区别的是Tognini-Bonelli(2001),她以any一词的用法为例阐

释了基于语料库的研究范式。传统的语法理论认为, any一词通常用于否定句中。Tognini-Bonelli在语料库中发现了丰富的使用实例, 充分证实了这一点。但她同时也指出, any一词不仅用于否定句中, 而且还有许多其他的用法。在基于语料库的研究范式中, 语料库的作用是试金石, 无法展现原有理论和观点中并没有提及的现象。梁茂成(2016: 34)对此评价道: “在这种研究范式中, 研究者已经对理论有足够的信心, 语料库在其中并不起决定性作用, 充其量也只是对原有理论进行局部修正。”

同时, 梁茂成(2016)认为基于语料库的研究范式与普通的实证研究十分相似, 不同的是, 这种研究范式所使用的数据并非问卷、访谈或其他诱发数据, 而是自然发生的语言。除此之外, 二者的研究过程也十分相似: 都是从现有某种理论出发, 提出一种理论或结果假设, 然后到语料库中检索, 观察语料库中的语言使用实例是否支持上述假设。当然, 作为实证研究, 研究者通常还需进行某种统计分析, 比如差异检验(如语料库分析中的对数似然比和卡方检验)、相关性分析以及更复杂的多元回归分析和聚类分析等。

当前的语言研究很大程度上都是在已有的理论框架下展开的, 研究者们并不需要很多专门的训练就可以操作基于语料库的研究范式。正如前面所说, 因为这种研究范式与常见的实证研究方法十分相似, 所以可以说, 基于语料库的研究范式似乎更能够为学科背景各不相同的学术英语研究者广泛接受。

1.2.2 语料库驱动研究范式

语料库驱动(corpus-driven)的研究范式是以Sinclair为代表的伯明翰大学语料库语言学研究者主张的, 它采取的是一种源自文本证据的“自下而上”(bottom-up)的研究路径(Francis 1993)。语言研究者根据语料库所提供的语言实例, 对该语言进行全面的枚举和描写。在描述过程中, 研究者不会刻意修改、忽略或者摘取一部分证据来迁就已有的理论, 从而

充分反映语言事实。例如, Francis (1993) 观察了 possible 一词在语料库中的使用实例, 描写了该词所有的使用环境, 并找出了几个重要的且被语法理论所遗漏的结构。其一是“it+be 动词+possible+that/to”结构。她通过观察所有的语言实例, 总结出两种语法情形: that 从句表达可能性(如 It is possible that the zoo will have to close.) 和 to 从句表达能力(如 It is possible to hire towing cars and trailers if you require them.); 其二是“定冠词 the+形容词最高级+possible+中心名词”。possible 既可以置于中心名词之前, 也可以置于中心名词之后, 而且这种句法结构是 possible 所特有的; 其三是“as+形容词+as+possible”结构, 该句法结构也是 possible 所特有的。由此可见, 语料库驱动的研究是从词汇入手, 采用自下而上的方式观察语言的句法结构和使用环境, 通过描写所有语言使用实例, 补充或修正现有的语法理论和观点。对此, Francis (1993: 147-148) 认为, “这类观察的概括性能够质疑语言的现有观点, 因为词汇的语法情况是多变的, 常常跨越单一的既定规则。语料库驱动的研究经由自下而上的语言范例, 修订和更新各类语法规则和类别”。

Francis (1993) 将语料库驱动的研究脉络大体概括为由使用实例到语言词项, 再到语法语境。Tognini-Bonelli 将其进一步归纳为“是由观察到假设到概括再到统一的理论陈述 (observation leads to hypothesis leads to generalisation leads to unification in theoretical statement)” (Tognini-Bonelli 2001: 85)。同时, 她强调“这一研究方法并不是机械的, 而是由作为语言学家的研究者利用自身的知识、技能和智慧不断地进行调整和修正” (Tognini-Bonelli 2001: 85), 直到最终找到能够概括所有语言事实的理论和框架。可见, 整个过程并非完全受语料库驱动, 而是研究者面对大量的语言事实, 发挥主观能动性, 经过观察形成假设, 再观察更多的语言事实, 对假设进行修正和调整, 直至最后的理论概括和统一(梁茂成 2016)。正如 Tognini-Bonelli (2001: 85) 所言, 在语料库驱动的研究中, 没有单纯的直线式归纳。可见, 采用语料库驱动的研究范式, 研究者观察和描写语料库

中的语言实例，自下而上地归纳和概括语言现象，从而通过对索引行中反复出现的语言规律及其频率分布进行系统归纳和分类，得到不同于以往研究的语言学范畴(梁茂成 2016: 33-34)。

在学术英语方面，语料库驱动研究比基于语料库研究少，原因之一或许正如之前所谈论到的，学术英语研究以需求为出发点，研究者或者学术英语教师通常以自己感兴趣的某个学术语篇语言特征作为出发点，考察该语言特征在语料库学术文本中的频率、形式和功能。但是，我们需要更多的语料库驱动的方法观察学术词项的语言实例，从而丰富和发展学术英语语言特征的知识 and 理论。例如，Granger & Paquot(2009)关注学术语篇中实义动词的使用特点，研究将BNC的学术体裁作为专家语料库，将国际英语学习者语料库(ICLE)作为学习者语料库，通过VV or VV*标注检索实义动词，她们观察所有实义动词实例，对比专家语料库和学习者语料库，发现conclude和argue在学生语篇中的使用较多。通过进一步观察，她们发现在学生语篇中，conclude的突出使用情形是作为句首状语To conclude，而argue的突出使用情形是people argue和I argue。以上两个实义动词的惯用语法形式是以往语法规则和理论中没有涉及的。

1.3 学术英语界说与发展阶段

1.3.1 学术英语的概念

学术英语，泛指在学术语境下开展各类学习和交流所需的英语。在专业教材、学术专著、课堂用语、口头汇报、期刊文章和学位论文等一系列学术体裁中使用的英语都可称为学术英语(Hyland 2006; 熊淑慧、邹为诚 2012)。根据Hutchinson & Waters(1987)提出的英语教学体系(见图1.1)，学术英语和行业英语(English for Occupational Purposes, 简称EOP)同属于特殊用途英语(English for Specific Purposes, 简称ESP)，

常与特定学科和专业结合在一起，如自然科学、生命科学和社会科学等。Jordan (1997) 曾将学术英语分为通用学术英语 (English for General Academic Purposes, 简称EGAP) 和特殊学术英语 (English for Specific Academic Purposes, 简称ESAP)。前者指不同学科专业通用的英语和话语实践共核, 后者是学生在特定的学科与专业, 完成知识学习与交流所使用的英语。但是, 通过上述体系, 我们看到学术英语是与特定学科和专业相关的, 因此在现实中几乎不存在通用学术英语文本。Hyland & Hamp-Lyons (2002) 强调, 学术英语受特殊用途英语理念影响, 在理论和教学上应突出特殊用途而不是通用用途。卫乃兴 (2016: 270-271) 认为, 学术英语本质上是英语的一个特殊子类, 话语事件参与者是具有专业知识的学术人员, 话语行为涉及专门的词汇和语法以及意义与功能, 非话语行为涉及专用公式、图表及其他元语言, 话语效果涉及专业知识的传播和信息的交流。

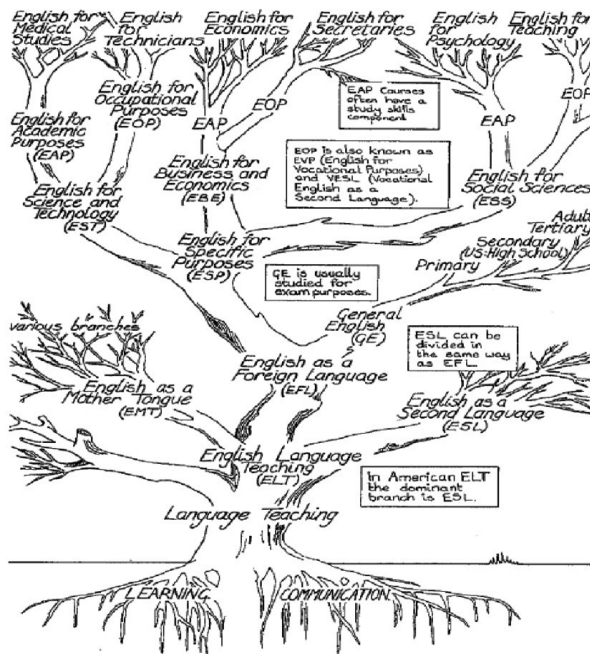


图 1.1 英语教学体系 (Hutchinson & Waters 1987)

学术英语不仅是英语教育的一种倾向,适应和满足学生需求,同时也是应用语言学一个重要的研究方向。它涉及学术语言的需求分析、文本分析、话语活动分析以及教学与测评等议题(Hamp-Lyons 2011; Hyland 2009)。一般来说,学术英语始于分析学生对学术语言的需求,止于对文本和话语实践的解析,最终反哺课堂教学和测评。例如,学生用英语撰写学位论文时存在诸多困难,需要增强把握该体裁的能力,因此学术英语教师需要根据该需求,了解学位论文的目标读者群和话语共同体,分析该体裁的语言和修辞特征,进而安排教学内容,提高学生驾驭该体裁的体裁意识和话语能力,构建学术立场和身份。可见,学术英语在数据分析基础上为学生提供知晓专业领域和学科“文本准则”(textual norms)的有效“路径”(access)(Hyland 2018: 383-384)。所以学术英语是外语教学与研究的一种取向,它突出目标话语共同体,以文本为切入点,通过实证描述目标话语共同体共享的特定语言特征、体裁资源和话语实践,并以此组织教学,提高学生体裁与修辞意识,使其了解交际目的和学科文化,从而更好地加入目标共同体,构建学术身份。

需要强调的是,学术英语与其他类别话语的分析不同,突出之处在于学术英语的研究与教学紧密结合,其理念是通过挖掘和描述学术话语的语言特征,指导课堂教学,增强其针对性与有效性。我们通过了解与学术英语密切相关的两种国际刊物 *Journal of English for Academic Purposes* 与 *English for Specific Purposes* 的宗旨和兴趣,也可以发现学术英语研究的教学落脚点,即对学业和学术交流语境下的英语使用给予语言学描述,用于特定共同体话语实践的教学。

1.3.2 学术英语的特征

通过上述对学术英语的界定可以看到,学术英语教学与研究的出发点是解析学术英语的语言本体特征。Snow & Uccelli (2009) 认为学术英语具有在英语语言中融入偏学术内容的特点。学术英语的读者群体是学术话

语共同体中的学者，他们共同秉持特定的语言使用和话语形式。例如，学术期刊论文和大众科普读物虽然都是英语文本，但是面向的读者不同，使用的语言特征与交际倾向也有所不同。大众科普读物的阅读人群来源广泛且背景各异，为了增强文本的广泛可读性，语言用词较为直白，语法结构简单清晰。相比之下，学术期刊论文的读者是学科领域的专业人士，他们熟悉且惯用一套约定俗成的话语方式来表述信息和撰写论文，因此该语言通常包含信息密度高的词汇形式以及大量的复杂句式。

Snow & Uccelli (2009: 119) 列举了学术英语在人际立场、信息容量和信息组织等方面的语言特征 (见表 1.1)。

表 1.1 学术英语的语言特征 (Snow & Uccelli 2009: 119)

1. 人际立场 无偏见的、权威性的学术立场
2. 信息容量 信息密集且内容简洁
3. 信息组织 结构衔接，明示关系，连贯逻辑，不言自明
4. 词汇选择 词汇多变，正式表达方式，词义准确，抽象/术语概念
5. 信息呈现 一致且复杂，或非一致紧凑；抽象概念作主语
6. 推理策略 劝谏性语步组合

可见，学术英语和日常英语截然不同。日常交际语言的态度立场多是热情洋溢、情感丰富的，信息组织以独立小句为主，用词也基本上是口语化和语义模糊的词汇 (如 sort of、something like)。对于学术英语写作的

语言特征, Hyland (2017: 26) 将其概括为八个方面: (1) 明示篇章结构与交际目的; (2) 依靠引用支持论点; (3) 强调行动本身而非行动的施动者; (4) 限用修辞性问题; (5) 少有话外音; (6) 谨慎持言, 慎重立论; (7) 多表事物, 少表事件; (8) 列清推论步骤, 明晰句际衔接。下面将结合示例的学术论文引言片段, 分别论述这八个特点。

The measurement of complex permittivity with respect to frequency—dielectric characterization—has found use in diverse applications in industry and throughout the biological, chemical, and pharmaceutical sciences [1]-[6]. However, it remains a specialist technique that has not been widely adopted, largely due to the popularity of established techniques such as UV-visible absorption, infra-red, and mass spectrometry. This is despite its great potential for nondestructive, noninvasive, non-contact, and label-free measurements of multi-phase liquid systems [4]-[9]. This paper describes modifications of the split-ring resonator (SRR) that enable high resolution, low cost, and easily integrable sensors for performing such measurements. (Rowe *et al.* 2014)

首先, 该段落呈现明显的篇章结构, 先表明相关领域, 再指出领域的研究缺口, 进而表明当前研究旨在填补当前研究缺口。该篇章布局不仅文脉清晰, 而且具有说服力, 充分反映出作者引入当前研究并吸引读者的交际意图。具体来说, 首句回顾前人研究, 点明相关领域。第二句试图找出前人研究中未涉及或者挖掘尚浅的方面。第四句表明文章旨在填补该空白的研究目标。

其次, 在该引言段落中, 文章引用大量的文献来支持论点, 例如第三句。引用文献在支撑观点的同时, 也提供了当前研究的文献关联和学科理据。

第三，文章突出研究的行动本身，而非行动的施动者。例如：在 measurement of complex permittivity and a specialist technique that has not been widely adopted 中，作者突显行动“测量”（measure）和“采用”（adopt），淡化谁进行测量和谁未能采用。

第四，我们在该段引言中还看到，文章一直使用陈述句，避开修辞性疑问句。

第五，该段引言没有赘余的话外音，贯穿始终的都是研究议题。

第六，文章作者在提出观点时保持谨慎态度，为其所言所述留有话语空间和余地。例如：在第二句中，作者在描述原因时加入 largely 一词，为其原因推断留有话语空间和回旋余地。

第七，文章多以名词结构标明事物，少用动词勾画事件。例如作者使用名词形式 measurements of multi-phase liquid systems 和 modifications of the split-ring resonator，而非动词 measure 和 modify，以此强调事件的事实属性。

第八，文章通过 however 点明现有研究的不足，再通过 this paper describes 体现文章的研究目的。此外，文章利用连接代词结构 this is、such measurements 明示句子之间的衔接关系。

但是，学术英语并不是完全剥离于日常语言的，而是日常语言、抽象语言和隐喻语言的结合（Fang 2016）。它们之间的关系如图 1.2 所示，同时表 1.2 分别举例描述日常语言、抽象语言和隐喻语言的不同特征。

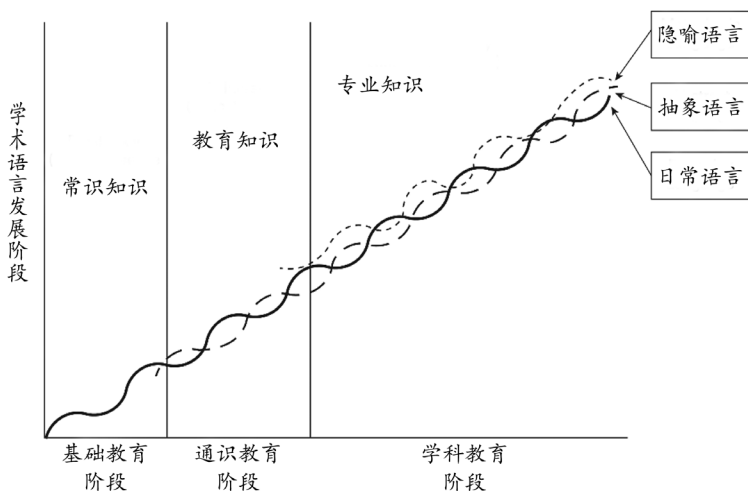


图 1.2 学术语言与日常语言、抽象语言和隐喻语言的结合

(改编于 Fang 2016: 326)

表 1.2 日常语言、抽象语言和隐喻语言的举例与特征 (Fang 2016: 325)

语言形式	举例	特征
日常语言	You can control the trains this way, and if you do that, you can be quite sure that they'll run more safely and more quickly, no matter how bad the weather gets.	<ul style="list-style-type: none"> 句式由并列句和从属句构成 (and、if、no matter how) 过程为导向 (control、do、run、be、gets) 代词为主语 (you、they) 口语词汇及短语
抽象语言	If this method of control is used, trains will unquestionably be able to run more safely and faster, even when the weather conditions are most adverse.	<ul style="list-style-type: none"> 句式以从属句为主 (if、when) 事物或概念为导向 (this method of control、trains、the weather conditions) 抽象物体/概念 (this method of control、the weather conditions) 或具体事物 (trains) 为主语 正式的词汇和短语

(待续)

(续表)

语言形式	举例	特征
隐喻语言	The use of this method of controlling unquestionably leads to safer and faster train running in the most adverse weather conditions.	<ul style="list-style-type: none"> 句式通常由动词词组 (leads to) 连接两个较长的名词短语 (the use of this method of controlling、safer and faster train running in the most adverse weather conditions) 紧缩式名词短语表达抽象事物/概念 (the use of this method of controlling、safer and faster train running in the most adverse weather conditions) 正式的词汇和短语

由图 1.2 可见, 学术语言在日常语言、抽象语言和隐喻语言中的具体倾向程度不仅取决于特定的学习阶段, 也与学科知识内容有关。在基础教育阶段, 学习内容往往是常识, 因此语言偏向于日常语言。在通识教育阶段, 学业内容以高校通才教育知识为主, 涉及的学术语言多是抽象语言, 用以描述和传授关于知识世界的一般内容。在学科专业教育阶段, 绝大多数的学习内容是与学科领域相关的专业知识, 所使用的学术语言是伴有语法隐喻(如名物化)的语言。但是另一方面, 学术语言在日常语言、抽象语言和隐喻语言中的倾向程度也和语篇表达的知识信息特点有关。Halliday & Martin (1993) 曾指出, 作为科学知识传播的载体, 学术语言的具体表现形式随着学科和专业的不同而变化。譬如, 自然科学文本主要报告技术信息与科学发现, 知识内容的专业性和技术性较强, 因此学术语言倾向于使用以短语为主的隐喻语言。下例是来自材料学科的学术论文片段。很明显, 该语篇由名词短语结构组成: 一方面, 句子主语由名词表达的抽象事物构成, 如 sodium ion batteries (SIBs); 另一方面, 逻辑关系是通过短语结构建立 (with the abundance and low cost of sodium), 并不是由日常或抽象语言中常用的连词进行衔接(例如 because、since 等)。

Sodium ion batteries (SIBs) have been extensively studied in the recent past as an alternative to their lithium counterparts (LIBs), owing to concerns regarding cost and limited reserves of lithium. Particularly, in large-scale storage applications like hybrid vehicles and renewable power stations, SIBs, with the abundance and low cost of sodium, are more viable than LIBs. (Banda *et al.* 2017)

相比之下，人文及社会科学语篇更倾向于抽象或者日常语言，因为此类学科文章论述的多是与社会和人们生活相关的实际问题。正如下例营销学论文所示，它包含日常语言，如以人作主语（individuals、they），夹杂着抽象语言，如以事物作主语（research）以及通过各类从句表达逻辑关系（where、whom）。

Individuals increasingly participate in virtual support communities (VSCs) where they conduct numerous aspects of their lives with others whom they may never encounter in person, and they interact within these communities to attain various goals. Research finds that individuals are more likely to achieve success with such goals when they make a public commitment to achieving them. (Bradford *et al.* 2017)

综上所述，学术英语复杂的语言特征促使其研究以文本为切入点，通过文本分析，找出目标语境和话语共同体的特定语言特征、话语实践与实际策略。随后以此组织教学，帮助学生了解和学习学科和专业的文本准则与话语规约。

1.3.3 学术英语发展阶段

学术英语起源于20世纪60年代（Hamp-Lyons 2011；Swales 2001）。

当时，英国文化教育协会 (British Council) 资助大量发展中国家学生赴英国学习。随着短期培训课程的增加，学者们开始探讨英语教学出现的新需求，包括学业学习与学术交流等所需要的英语能力。Swales (1985) 提到，学术英语相关研究的雏形始于 Charles Barber 在 1962 年发表的文章 “Some measurable characteristics of modern scientific prose”，并大致经历了萌芽、成长、突破和成熟四个重要的发展阶段。

20 世纪 60 年代是学术英语的萌芽阶段。当时，一些研究 (如 Halliday *et al.* 1964) 针对传统英语教学出现的新需求 (如为印度发电站工程师讲授英语)，提出语言学的任务是基于特定人群和特定语境下的语言实例，分析包括学术交流在内服务于特殊用途的语言特征。但是，当时正值自然科学与社会科学追求客观主义和实验方法的时期，学术英语被认为是简单直白的语言符号，是对知识信息的机械编译。因此，那个时期的学术英语研究仅停留在词汇和句式浅层的统计分析上，例如 Barber (1962) 统计三篇科技文本的动词形式，发现 28% 的实义动词是用于被动语态，而被动语态的最常见时态是现在时时态，约占各时态总数的 25%。

学术英语在 20 世纪 70 年代得到进一步发展。首先，学者们开始观察口语语篇 (如课堂话语)。第二，研究逐渐跨越词汇和句式简单的统计分析，开始关注语篇模式和规律，其中 Halliday 和 Hasan 于 1976 年出版的 *Cohesion in English* (《英语中的衔接》) 是较有代表性的成果。第三，学术文本分析开始采取修辞结构视角 (如 Selinker *et al.* 1972)，文本被分为开端、中间和结尾阶段式的修辞结构，而且学者们也逐渐认识到文本具有实现劝说读者的修辞功能。第四，教师逐步加强对语言使用规则和修辞策略的教学。Swales (2001) 评价道，上述四个方面的进步为学术英语的后期发展奠定了重要的基础，进而摆脱形式语言学的影响，吸纳多学科的理论 and 视角。

20 世纪 80 年代是学术英语取得突破性进展的关键时期，这一时期产生了三个重要概念：体裁、科学知识社会学以及互文性。Tarone *et al.* (1981) 和 Swales (1981) 首次将体裁 (genre) 一词应用于学术英语文献。

体裁概念逐渐发展为精细分析语言选择的系统功能语言学学派、强调交际语境的新修辞学派和折中语言与语境的特殊用途英语学派 (Hyon 1996)。科学知识社会学为学术英语的发展奠定了重要的阐释学视角。其观点认为,科学事实建构在社会活动和修辞实践的基础上 (Knorr-Cetina 1981), 强调学术话语研究者在学术活动和知识建构中的作用。此外, 在这一时期, 互文性 (intertextuality) 对学术英语的研究与教学也产生了较大的影响。互文性指任何语篇都是对先前话语的回应, 同时也反作用于后续的话语文本, 突出了语言在各类学术话语活动中的重要性 (Bakhtin 1986)。

20世纪90年代至今, 学术英语从此前三十多年的词汇语法分析, 逐渐进入到多视角、多维度的成熟发展时期。在这一时期, 学术英语产生了语步、学科间性和修辞手段三个关键词。首先, 在Swales (1981, 1990) 体裁分析的基础上, 大量研究剖析了学术体裁不同语步的交际目的及其语言实现形式。其次, 研究聚焦学术英语的体裁间性 (如Hyland 2000; Hyland & Bondi 2006), 比较不同学科在话语实践和知识建构方面的相同点与不同点。第三, 与20世纪70、80年代的学术英语语言分析不同的是, 研究者不再聚焦时态和语态等词汇句法的表面特征, 而是关注学术话语的修辞手段和语言策略, 比如元话语、词块和立场等语言资源。

1.4 学术英语的主要研究路径

学术英语研究的文本传统侧重语篇的语言选择、论辩规律和意义功能 (Hyland 2009), 因此形成了以系统功能语法、体裁理论和语料库为主的知识基础与研究路径 (Charles & Pecorari 2015; Ding & Bruce 2017)。

1.4.1 系统功能语法

系统功能语法作为一种功能学派, 强调语言选择的使用情境及语篇功

能。语言被认为是社会符号系统，语言的选择实现和表达特定的功能与意义，是系统和功能的结合。系统是指在特定社会语境下对词汇语法体系的语言选择，功能是语言系统选择所实现的社会意义。在系统功能语言学理论中，语言系统的选择实现三个元功能，即概念功能 (ideational metafunction)、人际功能 (interpersonal metafunction) 和篇章功能 (textual metafunction)。其中概念功能由经验 (experiential) 功能和逻辑 (logical) 功能组成，经验功能指用语言谈论世界经验和描述周围所发生的事件或情形，而逻辑功能是表示小句与小句之间关系的机制，通常以并列关系或从属关系的形式体现。人际功能表示建立语言和保持人际关系、影响他人的行为以及表达对世界的看法和态度。篇章功能刻画语言组织信息的方式，表明信息与信息之间的关系。

较早涉及学术英语研究的是分析科学文本中的主位推进模式 (Hutchins 1977)。主位推进模式一直是系统功能语法在学术语篇篇章功能研究中的主要应用，该模式试图从Hutchins所谓的“微观结构”出发，分析“宏观结构”的信息安排。例如，Nwogu & Bloor (1991) 考察医学语篇的主位推进规律，研究发现医学论文的正文和摘要与大众科普杂志中的医学报道在模式选择上存在差异。医学论文正文以“同一型主位模式” (constant theme pattern) 为主，大众科普杂志的医学报道以“线性主位模式” (linear theme pattern) 为主，而医学论文的摘要则介于两者之间。此外，Ventola & Mauranen (1991) 对学术语篇主位推进模式进行了跨语言的对比研究。研究通过分析芬兰学者的英文语篇，发现他们几乎毫无例外地使用“同一型主位模式”和“线性主位模式”，而英语本族语作者则使用各种不同的推进模式。研究还发现芬兰学者较少使用主位连接手段，且连接手段也比较单一，仅限于moreover、thus、therefore和also等少数几个连接词。

系统功能语法在学术英语研究中的另一个重要应用是基于“评估系统” (Martin & White 2005) 探讨语篇中的人际意义。评估系统关注

的是词汇语法层面以及篇章层面的评价语义,由态度(attitude)、介入(engagement)和级差(graduation)三个子系统组成。态度子系统包括情感(affect)、判断(judgement)和鉴赏(appreciation),分别指人们积极或消极的感情(如高兴、担忧、感兴趣、厌烦等)、人们对行为的态度(如赞美/批评、表扬/谴责等)以及对一些现象是否有价值的评估(如是否完善、美丽等)。介入子系统分为单声与多声评价,描述作者如何运用语言调节人际关系,传递主观判断,以此“结盟”相同观点者,“疏离”不同观点者。单声评价仅提供唯一立场,多声评价则暗示多种立场共存。对于其他立场,多声评价有收缩与扩展两种倾向。收缩评价反对其他立场,闭合对话空间;扩展评价包容其他立场,为读者预留选择立场的余地。借助级差子系统,情感、判断和鉴赏可用不同强度的语势表达作者态度,或者通过不同典型性与确切性的聚焦,介入别人的态度。Hood(2010)和刻璇、徐玉臣(2011)分析学术语篇中评估系统的语言实现形式,并列举作者是如何使用评估资源与不同的学术立场结盟的,进而构建与读者的人际关系。

1.4.2 体裁理论

体裁不仅指文本和交际事件的类型,而且是“一种有步骤的、以交际目的为导向的社会互动过程”(Martin 1992: 505)。体裁联结篇章中的语言表达形式和交际目的,通过交际目的来解释特定体裁的篇章结构和语言特征。前面谈到,Hyon(1996)依据不同的研究侧重和教学倾向,将体裁研究归纳为新修辞、系统功能语言学和特殊用途英语三个学派。

新修辞学派指北美学者在研究一语写作(composition)和修辞时涉及的体裁视角。它着重关注体裁发生的语境及其社会行为与目的。正如Miller(1984: 151)在该学派开创性文章“Genre as social action”中指出的那样,“体裁的修辞意义不是围绕其语言内容和语篇形式,而是关注它所实现的社会实践”。因此,新修辞学派强调考察体裁发生和使用的社会

情境，通常运用民族志 (ethnography) 等方法深入描述体裁参与者共同体的态度、价值观和信念等主体认知。与学术英语相关的有影响力的文献是 Bazerman (1988) 和 Myers (1990) 两部著作。Bazerman (1988) 通过探索学术期刊论文百余年的历时演变，发现学术论文文本表征与科学共同体的学术发表实践、科学范式以及普世价值等社会文化变化息息相关，认为该体裁作为重要的知识载体推动了学术语境的发展。Myers (1990) 以期刊手稿和项目申请书为焦点，关注年轻的生物学家是如何通过这些学术体裁跻身学科共同体的竞技场，获取学界的认可与支持的。可见，该学派学者强调体裁的动态发展属性，即体裁随着作者与读者交际意图的改变而变化，或者说，体裁演变也促使参与者的交际目的发生改变。因此，新修辞学派学者提倡重视体裁对语境和参与者的创造潜能，认为体裁研究应该在互文中进行，了解参与者是如何根据对体裁的先验知识而展开社会行为的。但是，该学派并不关切二语课堂教学，同时认为课堂并非真实的学术环境，因为它缺乏真正的学术互动与读者交际 (Hyon 1996)。

相较于新修辞学派，系统功能语言学学派和特殊用途英语学派一方面关注教学需求，另一方面侧重体裁的语言本身，探究交际功能和语言形式之间的关系。系统功能语言学学派受其语言学理论溯源的影响，旨在深入分析情景语境中语言形式和功能之间的联系。该学派认为，情景语境由语场 (field, 指谈论的内容)、语旨 (tenor, 指参与者的身份及相互之间的关系) 和语式 (mode, 指话语的修辞方式和交际功能) 构成，体裁的语言选择与情景语境有着系统性的关联 (详见 Halliday & Matthiessen 2013)。例如，手册和指南这类引导类体裁包含大量的行为过程小句，因为该类体裁突出执行操作的语场；教材中通常含有高频的关系过程小句，因为说教类体裁主要传递事实信息的语场内容。语场、语旨和语式又被称为语域变量，因此系统功能语言学学派认为体裁是“语域+目的”，是以特定的语言结构来配置语域资源，从而达到交际者的交际目的。此外，系统功能语言学学派还关注语篇的图示结构 (schematic structure)，分析文本结构和

句子层面的语言选择,如词汇、语法和衔接手段等。例如, Coffin(2006)通过考察“历史描述”(historical account)类学习者体裁的图示结构,发现该体裁由背景、描述序列和推论三个阶段构成。在描述序列阶段,学习者按照时间顺序陈列事件,但并不是简单地一一讲述事件,而是以前一事件作为后续事件承担施事功能的形式展开。这一体裁图示结构及其交际功能致使描述者运用大量名物化的语法隐喻,在句首主位使用名物化结构表达施事动作。

特殊用途英语学派虽然也受教学驱动影响,但是有别于系统功能语言学学派。前者旨在帮助英语非母语的学习者掌握学术和行业英语的语言特征,而后者以中小学移民的弱势学生群体为对象。此外,两个学派对体裁语言分析的精细程度不同。如果我们设定“语言”和“语境”为一个连续体的两个终端,那么系统功能语言学学派位于最左侧,新修辞学派在最右侧,特殊用途英语学派则位于中间位置。Hyland(2002: 115)评价道,“特殊用途英语学派采取折中视角,既借鉴巴赫金的互文性和对话性概念,又融入系统功能语言学的文本解读思想。它是系统功能语言学的应用,但却并不倚重语法分析的层级、系统性以及纯理功能的示例与体现”。该学派认为,体裁是一类具有相同交际目的的交际事件,话语共同体成员普遍接受和认可该交际目的,同时交际目的又决定体裁的存在,使得同类体裁语篇具有大体相同的图示结构,并影响文本内容和语言选择(Swales 1990: 58)。因此,该学派的体裁研究主要由两部分组成:一是描述语篇组织结构的语步分析;另一个是句子层面的词汇、语法特征考察。语步分析最著名的当属Swales(1990)提出的“创建研究空间”(Create A Research Space,即CARS模型)模式。Swales基于158篇多学科的科研论文引言部分语料库,发现这些学术论文在组织结构方面有规律可循,且相当一致,基本上都遵循CARS模式。他将论文引言结构概括为三个语步:语步一,确立研究领域(establishing a territory),通过回顾前期研究成果阐述研究问题的重要性,其中三个步骤包括建立中心议题,概括

论题内容及回顾前期研究成果，其中第三个步骤为必要步骤；语步二，确立研究缺口 (establishing a niche)，指出前期研究的不足，再次重申研究的必要性，其步骤包括反驳已有观点，指出不足或空白，提出问题或继承前期研究成果，其中步骤二和步骤三是必要步骤；语步三，弥补研究缺口 (occupying the niche)，告诉读者引言中提出的问题将在正文中如何解决，其中又包括概述研究目的，综述当前研究状况，通报主要发现以及介绍科研论文结构等步骤。随后，Swales认为在批量的文本基础上观察总结的体裁组织结构和语言使用是“反复的、具有规律性特征的” (repetitive regularity)，而这种规律性语言特征是“不同研究领域进行学术交际的最好符号” (Swales 1990: 171)。在发现语步结构的基础上，Swales (1990: 144) 找出了实现语步交际目的的词汇语法特征，譬如在确立研究领域时惯用的表达方式。现在完成时是典型的语法特征，如 *recently, there has been wide interest in...* 和 *many investigators have recently turned to...* 等。这些都是被受众的话语共同体所接受和认可的语步结构与语言特征。

系统功能语言学和特殊用途英语学派倾向在语言学视角下，运用功能语法和语篇分析等方法，考察体裁交际功能的词汇、语法表达和修辞策略。相比之下，新修辞学派更多关注体裁的社会语境，而淡化词汇和语法的精细分析。这一学派常常使用民族志的方法，分析体裁的社会行为以及使用者共同体的态度、认知、价值观和社会行为。尽管 Johns (2003: 206) 认为“特殊用途英语学派的体裁视角越来越关注语境，与新修辞学派产生了更多的交集”，但是 Ding & Bruce (2017: 72) 认为，学术英语研究受系统功能语言学和特殊用途英语学派的影响最大，同时也借助语料库语言学的方法，大批量地识别了语步结构的词汇、语法形式与规律。

1.4.3 语料库

作为大规模自然发生的语言实例集合，语料库能够帮助我们观察到词

汇、语法和语篇等不同层次的语言特征在学术语境中反复使用的规律。下一节将详细阐述语料库与学术英语研究的结合。

1.5 语料库与学术英语的结合

前文谈到, Charles Barber 在1962年发表的文章“Some measurable characteristics of modern scientific prose”标志着学术英语研究的开端。Barber 基于2.3万余词的三篇科技文本,统计了动词的时态和语态形式,发现28%的实义动词用于被动语态,被动语态的最常用时态是现在时,约占25%。可见,学术英语研究自始便与语料文本和语言观察结合在一起,并受益于语料库。正如Hyland(2012a: 30)所言,“很难想到比学术英语研究受语料库影响更大的应用语言学研究”。

学术英语在观察和分析学术活动与话语实践的基础上组织教学,满足学生学科专业学习和各类学术交流的需求。可见,学术英语研究“以解析学术交流涉及的语言(包括词汇、语法)、话语和体裁而展开”(Dudley-Evans & St John 1998: 5)。了解这些语言、话语和体裁信息最为有效的途径莫过于接触话语活动和语篇文本的实例。语言实例不仅为学术英语本体研究提供充分的实据材料,同时也为学术英语教学提供参考学习范例。语料库语言学以真实的语言事实作为数据来源和研究对象,为学术英语研究提供了实例文本库和全新的研究思路。语料库语言学基于丰富的真实语言资料,从调查语言信息的分布频率入手来研究语言,以实际使用中的语言现象的出现概率为依据进行词汇或语法形式的统计分析,从中寻找语言使用的规律,进而对语言、语言交际和语言学习的行为规律进行多层面和全方位的研究(徐秀玲、许家金 2017; 杨惠中 2002)。Nesi评价道,“毫不夸张地说,所有的语料库研究路径都有益于学术英语,它的优势就在于几乎每一次对学术语料库的检索都会给学术英语研究者提供未知的新信

息……语料库有助于提高人们对学术语篇的种种认识，包括词汇、语法、短语和体裁类别等”（Nesi 2016: 206）。

语言本体特征、教学发展与教材开发是学术英语研究的三大议题，即揭示学术英语的语言特征，提高学术英语教学实践，设计教学材料和资源。根据相关学术英语综述研究（孙云波、冯婕 2014: 93），几乎95%的语言本体研究都基于语料库，涵盖学术英语语法、语体、词汇、语篇和体裁等诸多方面。语料库观察和分析建立在大量真实文本和自然发生的语言使用基础之上，能够最大限度地获得语言特征普遍化的归纳信息，而这种归纳信息一方面概括学术语言的一般特征，另一方面增强研究结果的效度与可比性。杨冬玲、陈坚林（2015）针对学术英语研究进行知识图谱演化趋势分析，发现在2001年至2005年间，应用语料库手段的论文引用率高且覆盖面大。

尽管综述研究指出了学术英语和语料库结合的重要趋势，Ding & Bruce（2017）指出，评价一个研究方法是否适用于学术英语研究，需要考虑三方面问题：一是它会给学术英语研究提供什么信息；二是它能够如何提高学术英语范畴内的文本能力、体裁意识和社会语境知识；三是它涉及的语言学习理论能够如何促进学术英语能力的提升。前面两个小节指出，学术英语研究本着语言研究和教学的双重目的，一方面揭示学术语言的本体特征，另一方面结合语言本体分析反哺语言教学。学术英语研究强调在获得学术语言特征的基础上增强学习者对该语言特征的了解，进而提高学生参与学术对话与知识交流的话语能力。Bhatia（2004: 144-145）认为，话语能力包含文本能力、体裁意识和社会语境知识。其中，文本能力不仅包括准确掌握词汇和句式等浅层语言特征的本领，还涵盖驾驭文本、修辞和语用知识解构与重塑文本的本领；体裁意识指学生能够识别、建构、阐释及充分运用各类体裁及要素资源，完成话语实践，融入话语共同体的能力；社会语境知识表示可以运用语言参与社会与机构活动，表达社会身份的素质和能力。由此可见，Ding & Bruce（2017）提出

的上述三个问题主要是围绕学术话语能力展开的。接下来,笔者将分别探讨这三个问题,进而评判语料库对学术英语研究的价值和适用性。

第一,语料库给学术英语研究提供什么信息?

词表(wordlist)、主题词(keyword)和索引(concordance)是语料库处理文本样本的三个主要手段(梁茂成等 2010)。

首先,词表确切来说指的是词频表,是一个语料库中所有词及其出现频数的列表。它可以按词语出现的频数、词语占全部文本的百分比排列,也可以按字母顺序排列。词频表可用来调查语篇词汇的密度、多样性、复现率、高频实义词等词汇类型或频数特征,进而比较某些词在不同文本或不同语体中的分布状态等。学术英语中较有代表性的词表研究是Coxhead(2000)的学术英语词表。它对涵盖学术文章、大学教材和实验手册的350万词学术书面语语料库展开词表研究。在排除英语最常见的2,000个词汇后,Coxhead分析了剩余词汇的词频和分布率,得出学术英语词表。该词表涵盖570个词族(word family),占学术文本全部词汇的10%,而这570个词族在同等规模的小说语料库中仅占1.4%。可见,该词表充分呈现出学术用途英语词汇的特征。Coxhead(2000)认为,从另一个角度讲,词表所呈现的词汇恰恰也是学生在各类学术交际中需要掌握的词汇,因此学术英语词表为学生和教师列明学生在学业知识学习以及教师在学术英语教学中需要重视的词汇。杨冬玲、陈坚林(2015: 3)认为,Coxhead(2000)在应用语料库方法产出学术词汇列表的研究领域具有里程碑的意义。因为在2000年之前,结合语料库与学术英语的研究数量有限,且语料库规模较小,主要使用手工统计的方式。由此可见,词表可以提供语料库词频方面的统计信息。

其次,主题词表是指某篇或某一批语篇中与参照语料库相比具有超高复现频率的词汇列表。主题词表可以用来揭示语篇的内容大意、文体风格、主题词之间的语法和语义联系,也可以揭示主题的情节发展。通过观察主题词,还可以发现某一即定文类或主题文本的词语特征。梁茂成

等(2010: 86)认为,“词语像人类一样聚群……一些特定的词语群总是被触发,表达相同或相似的主题”。根据这一特点,我们可以计算某一文本或某一类文本的词语分布,并与其他文本的词语分布进行对比,把文本中那些显著高频的词语析取出来,生成一个词表。该词表可被看作是主题词表,反应文本的“所言之事”(aboutness)(Scott 2000)。生成主题词表需要两个语料库,一个是用于观察和分析的语料库,称为“观察语料库”(observed corpus),另一个是用于对比的语料库,称为“参照语料库”(reference corpus)。观察语料库可以是单篇文本,也可以是同类题目的多个文本。参照语料库可选用通用语料库,其总量要大于观察语料库。Römer & Wulff(2010)以密歇根大学高水平学生论文语料库(Michigan Corpus of Upper-Level Student Papers,简称MICUSP)为研究对象,以生物学分子库为观察语料库,MICUSP为参照语料库,得出生物学分子库的关键词表。species、gene、plague、cells、protein是主题性(keyness value)最高的五个词。这表明生物学文本涵盖的内容和学科特性。可见,主题词表提供学术英语中针对特定学科和体裁的特色词汇,是学生在专业学科知识学习中需要掌握的词汇。

索引又被称为“语境中的关键词”(Key Word In Context,简称KWIC),它是某一检索词在被检索的语料文本中出现的所有检索行的列表。通常,检索词居中,同时前后各有若干词,呈现其周边一定范围内的语境。索引可以按照检索词的字母顺序给出,或者检索词左右某位置的字母顺序列出,以便于我们从索引行中寻找规律。通过观察索引行中的检索词和它的“搭配”(collocation),我们可以了解该词的使用语境、规律和范式。Tribble(2002)以雷丁学术文本语料库(Reading Academic Text Corpus)的植物分子生物学文本为语料,检索activity并以字母顺序排列activity右侧的第一个单词,得到索引行如下(见图1.3):

urtains and let the sea breeze in before he got into bed
o leave this house . . . ' I broke in on her tirade. 'That's
a minute, Mrs. Landry,' he broke in gently. 'Loss of memor
act exactly the same,' she broke in. 'Please drive back. I
m the others is that my brother's in love with Lois. He nev
lking about when tea was brought in. Diana will soon be tw

图 1.5 LOB 语料库言情小说文本的 in 索引行

词表和主题词对教学大纲制定者和教学材料编写者来说具有重要的参考价值。教学大纲制定者可以根据词表和主题词拟定教学大纲的词汇部分；同样，教材编写者可以围绕词表和主题词进行教材编写，并辅以语料库索引设计相应的词汇和语法练习。

第二，语料库是如何增进学术英语范畴内的文本能力、体裁意识和社会语境知识的？

语料库研究通常涉及两个层面。首先是微观层面的词汇考察。上述词表、主题词等信息揭示的是学生在学术或特定专业学习语境中需要掌握的词汇和主题词。例如，Thompson (2006) 考察英国学术口语语料库 (British Academic Spoken English corpus, 简称BASE corpus) 经济学讲座语料，对其词表和主题词进行分析。他发现经济学讲座中仅有Coxhead (2000) 学术英语词汇表570个词族中的340个。此外，他通过主题词分析进一步发现，经济学讲座中名词的主题词特征尤为明显，高频名词主题词有：accounts、budget、capital、choice、commodity、constraint、cost、curve、debt、demand、elasticity、exports、income、liberalization、market、probabilities、profit、supply、trade、value、variable等，而let、represent、maximize、consume等则是常见的动词主题词。他认为，这些名词主题词体现经济学话语中“实体抽象化”和“过程具体化”的特点，而动词主题词的出现迎合了经济学话语的“符号表征”和“数理阐释”特征 (Thompson 2006: 263)。可见，语料库的

词汇发现可以反映出学生应该掌握的文本能力。正如Thompson(2006)的研究表明, 学生在学习经济学课程时, 不仅需要理解而且也要利用represent和let等主题动词建立符合表征和数理的关系, 营建经济学假说。

其次是语法句式层面。就语法句式而言, Charles(2000)以英语为母语的材料学和政治学博士论文为语料, 观察“引导型It”结构(例如, It is important to note...或It seems reasonable to attribute...)的频率和功能, 发现政治学论文使用“引导型It”结构的频率约为每十万词150例, 材料学论文的频率略高, 约为每十万词170例。在分析了该结构中谓语动词和评价形容词后, 她进一步发现两个学科论文通过“引导型It”结构表达不同的学术立场和声音, 并认为这是受学科知识特点和知识建构方式的影响。同时, Charles认为, 两个学科的论文在该结构使用上的相似之处源于一般意义上的学术行为过程和价值判断能力, 如合理、信度和效度等。因此, 该语法结构在一定程度上体现学术语篇的文本能力。

接下来的部分将通过学术引用(Hyland 1999a)和博士论文文献综述(Flowerdew & Forest 2009)两个实例, 讨论语料库在增强体裁意识方面的作用。文献引用(citation)在学术论文体裁中的知识建构与观点铺垫方面起着至关重要的作用。学术论文的重要交际目的是推介研究和观点。新知识的产生和其真实性架构在学科知识网络与文献载体上, 采取适当的引用方式和话语策略援引前人文献, 铺垫新知识产生的学科框架, 能使学科读者信服研究论文和结论观点。Hyland(1999a)通过语料库发现, 学术引用的主要形式是非嵌入式引用, 即把被引作者和文献置于引句之外, 从而突出被引文献, 淡化作者参与度。此外, 他还发现学术引用的另一个特征是以转述总结被引文献为主, 少有段落摘抄。在转述动词的使用方面, 理工学科论文以研究类动词为主, 如observe、show等, 而人文学科论文以言语类动词为主, 如suggest、argue等。可见, 语料库数据发现能够反映出学术论文体裁是如何将前人的观点纳入自己的论证的, 是如何评价和发展前人的观点的。

Flowerdew & Forest (2009) 以博士论文文献综述部分为语料, 以 *research* 一词为出发点, 发现 *research* 一词在文献综述部分的使用频率为 900 次, 且出现在 70% 以上的综述中。值得注意的是, 当论文作者在建立“揭示研究空白” (*indicating gap*) 语步时使用 *research* 一词的频率最高, 约每千词 8 次, 是其他综述语步的五倍, 而在“进行理论回顾和文献观点调查” (*surveying knowledge claims*) 时反而使用较少 *research*。他们考察 *research* 一词的搭配, 进一步发现两个惯用的搭配措辞: *little research has been done* 和 *further research is needed*。文献综述微观体裁的交际目的无外乎是找到文献缺口, 铺垫当前研究的价值和意义, 进而推动研究现状的发展。上述研究能让我们看到该词汇的使用频率和短语搭配是如何为体裁的交际目的服务的。如果把这种基于语料的发现展示给学生或者学术新手, 无疑会增强他们的体裁修辞意识和语用能力。

第三, 语料库涉及的语言学习理论能够如何促进学术英语能力的提升?

根据 Tribble (2002) 的观点, 学术英语能力由学科知识、写作过程知识、情境知识与语言系统知识组成。其中, 学生的自主学习主要围绕学科知识展开。尽管教师需要开展各类教学活动来帮助学生提高写作过程知识、情境知识和读者意识以及语言和体裁知识, 但是 Tribble (2002: 132) 认为, 在学术英语知识构成方面, 语料库是教师教学与学生自学的重要资源。通过以上分析与例证可以清楚看到, 语料库的数据发现为教师提供关于学术英语的文本知识、体裁和社会语境知识等方面的信息。

就语料库相关语言学习理论而言, 王龙吟、何安平 (2005) 和 Flowerdew (2015b) 认为“数据驱动学习”和“注意”是最重要的理论接口。

“数据驱动学习” (*Data-Driven Learning*, 简称 DDL) 由 Tim Johns 于 20 世纪 90 年代初提出, 是指在教室或课下自学时使用语料索引, 让学生探索目标语的使用规律。语料检索可以更好地体现以学习者为中心的发现式学习。例如, 当学生在撰写学术论文时无法确定 *data suggest* 和 *data*

suggests哪一个才是正确的语法使用的时候，便可以登录BNC网站，分别检索 data suggest 和 data suggests。学生会发现前者出现的频率为73次，而后者出现的频率仅为11次。约七倍的差距可以令学生清楚地看到 data suggest 更符合学术写作的常规，这将帮助学生决定在学术论文中使用前者。检索结果还将学术语境中使用 data suggest 的语境提供给学生，学生可以通过观察掌握 data suggest 的语用功能。这是DDL成功运用在学生自主学习中的例子。可见，它已经超越传统的课堂学习活动，成为一种学生独立自主进行发现式学习和探索式学习的模式。

由此可见，DDL是以语料库为基础，使用原始语料或通过语料库检索工具得到的结果进行语言学习的发现式和探索式学习模式。它可以是以教师为主导，将原始语料或通过语料库检索工具得到的结果编制成教学材料和练习。然后以学习者为中心设计一系列教学活动，引导学生发现目标语的规则，提高学生对有关知识点的敏感度。此外，它还可以是学生根据个人不同的学习需求和意愿，自己检索语料库，进行发现式和探索式的学习模式。根据语言习得的规律，语言知识的记忆不能保证语言使用的正确，必须有大量语言的输入才能使學生掌握所学习的语言。在语料驱动学习的过程中，语料库索引将大量的目标学术语呈现给学生，相比传统教学为解释某一学术英语规则在教材中提供有限的几个例子，DDL提供给学生的语言输入量要大得多。此外，由于语料库收集的是实际使用中的语言，它所提供的不仅是大量的语言输入，而且是真实的语言输入。与此形成对比的是，传统教材上的语言例子很多是由教材编写者根据自己的语言直觉编写出来的，并非真实语境下使用的语言。DDL所提供的语言输入在质和质量两个方面都优于传统教学中的语言输入。

就“注意”而言，Schmidt(1990)认为，它等于“注意力”与“意识”之和，它影响语言输入向语言纳入转化的过程。语言学习者在进一步处理语言输入或将其中的某种形式转化为纳入之前，必须注意到这种形式的存在并理解它，纳入就是学习者注意到的那部分输入。王龙吟和何安平认

为，由学生自己发现的凸显语言形式较容易被内化为纳入，“因为学习者必须首先经历‘注意’这个阶段，然后才能有所发现，这样一来，他们发现的语言形式必然也是他们注意到的语言形式，因而这样的凸显部分较容易被内化为纳入，进而有利于提升对该语言形式的意识”（王龙吟、何安平 2005: 30）。以下的教学实例体现了语料库索引对学生学术英语知识的提高作用。

在教师的指引下，学生分别以 *although* 和 *though* 为检索词，从鲁汶英语本族语论文语料库（Louvain Corpus of Native English Essays，也称 LOCNESS）中提取相关索引行。*although* 的索引行有 150 条，*though* 的索引行有 52 条。接下来，再观察 *although* 和 *though* 在句子中的位置并进行分类编号，进而归纳出四种语言形式，如表 1.3 所示。

表 1.3 *although* 和 *though* 的用法分类

类别 检索词	a 频率 (%)	b 频率 (%)	c 频率 (%)	d 频率 (%)	总频率
<i>although</i>	101 (67.3%)	39 (26%)	10 (6.7%)	0	150 (100%)
<i>though</i>	8 (15.4%)	9 (17.3%)	0	35 (67.3%)	52 (100%)

(1) 编号 a 代表 *although* 和 *though* 所引导的从句或短语在主句之前。

例句: Although laws may have changed, deep-rooted prejudiced attitudes remain in these adults.

(2) 编号 b 代表 *although* 和 *though* 引导的成分在主句之后。

例句: Sometimes they need to write just to express their pain or unsure feelings, though, more often, their purpose is to teach a lesson.

(3) 编号 c 代表 *although* 和 *though* 引导的成分在主语之后作插入语。

例句: Britain, although small, is more often than not in the fore of international matters.

(4) 编号d代表although和though后面不接任何从句或短语。

例句: He did nothing to help though and continued to walk by.

在although的150条检索结果中, 半数以上的although引导的从句或短语出现在主句之前(67.3%), 所有的although都用作连接词引导从句或短语, 无论它们在句中的位置如何。在though的52条检索结果中, though引导从句或短语出现在主句之前的情况只占总数的15.4%, 而在超过半数的情况下(67.3%), though不用作连接词。

王龙吟、何安平(2005)在上述教学实例基础上进行了评价: 学习者在教师的指引下, 利用语料库索引进行类似以上步骤的归纳, 可以发现一些语言形式方面的凸显特征。如上所述, 这些学术英语的语言特征是学习者自己注意到的, 因而更易于语言习得。

此外, 大量的实证研究表明了学生对学术英语课堂采用语料库方法和自行探索索引表现出了积极的态度(如Yoon & Hirvela 2004), 并肯定了语料库和数据驱动学习对提高词汇语法认识的作用, 增强了独立学术写作的自信度(如Yoon 2008)。就高级学习者而言, Starfield(2004)实地观察和跟踪记录博士生在进行二语研究学术写作时自助使用语料库的情况, 发现语料库的数据驱动学习有助于博士研究生策略性地选择使用学术英语资源, 从而建立适当的立场身份, 更好地融入话语共同体。可见, 语料库涉及的语言学习理论可以有效促进学术英语能力的提升。

总而言之, 围绕Ding & Bruce(2017)提出的三个问题, 我们在大量的实证研究和思辨阐释的基础上清楚地看到, 语料库语言学的语料、方法和理论充分契合学术英语的研究与教学。语料库通过提供语言使用频率和真实使用环境, 揭示学术英语特定的词汇和主题词以及基于学术体裁规约的语言使用规律; 另一方面, 数据驱动学习等与语料库相关的理论能够促进学术英语的学习以及学术英语能力的提高。

1.6 学术英语语料库类型

1.6.1 语料库的类型

常见的与学术英语相关的语料库包括自建语料库和开放语料库、通用语料库和专门语料库、口语语料库和笔语语料库、本族语者语料库和二语学习者语料库、专家语料库和新手语料库、共时语料库和历时语料库等，以下将逐一介绍。

第一，自建语料库和开放语料库。

根据学术英语研究的目标，研究者通常结合教学实践自建语料库。譬如，研究者对新兴的生物信息学学科的语篇特征感兴趣，便可搜集生物信息学领域权威的期刊论文，自建该学科语料库。又如，研究者对不同年级学生的学术论文语言特点感兴趣，便可搜集不同年级学生的学术论文，自建专门语料库。当然，我们也可以结合自己的研究目的选择已建成的开放语料库。例如，以不同年级学生学术论文的语言特点为研究目标，研究者除了选择自建语料库，还可以选用包括英国学术笔语语料库 (British Academic Written English corpus, 简称BAWE corpus) 和密歇根大学高水平学生论文语料库 (MICUSP) 在内的开放语料库。采用现成的开放语料库可以减少建设语料库的技术困难，简化搜集、清理、标注和赋码过程。在线的应用界面也为用户提供方便的操作方法。然而，自建语料库能够实现更为灵活、具体和专门的语料搜索和调查，这对于学术英语研究来说更具针对性。

第二，通用语料库和专门语料库。

通用语料库 (general corpus) 通常代表一种语言的整体，而专门语料库 (specialised corpus) 代表某一种用途或领域的语言。英国国家语料库 (British National Corpus, 简称BNC) 和美国当代英语语料库 (Corpus of Contemporary American English, 也称COCA) 是最为典型的通用语料库，它们分别代表英国英语和美国英语，既包含笔语又包含口语，同时涵盖各

类语言的使用场合和语域。因此，基于BNC和COCA的发现和结论往往是关于英国英语和美国英语整体的情况。但是，学术英语所用的通用语料库一般是代表多语域的学术英语整体，比如BNC学术语体集合。这类通用语料库可用于揭示学术英语有别于其他语体的专有语言特点。例如，Coxhead (2000) 集合学术英语各类体裁，包括期刊论文、教材、通用语料库的学术体裁部分和学术指南，整体考察了通用学术英语词汇的频率和覆盖率。

相对来说，学术英语中的专门语料库通常是代表某一学术体裁、某一学科和某一群体的语料库。比如，华中农业大学建立的农科医学英语论文语料库(刘萍等 2015) 是农科医学领域的权威期刊论文的专门语料库，代表该特定学科体裁的语言使用。对于学术英语语料库的建设和选择，语料涉及的学术语境越具体，体裁和学科类别越明确越好。

通用语料库虽然一般都含有学术英语部分。但是因为通常存在全文欠缺和体裁不均衡两方面问题，所以它们在学术英语研究中的适用性有限。这两个问题在学术英语研究中不可忽视。以BNC为例，BNC含有500个书面学术文本，约为1,600万词符，体裁涵盖论文、教材、专著、评论等。BNC的书面学术部分可以作为考察学术英语的整体语言特点的文本，譬如与BNC全库比较，分析学术英语的主题词。但是我们需要注意的是，BNC学术部分的体裁分布不均匀。在医学领域，90%的语料是期刊论文；在社会学和法学领域，60%来自专著，40%来自期刊论文；而在社会科学领域，80%来自专著，只有20%来自期刊论文。可见，我们应该谨慎使用BNC学术部分进行学术语篇的学科对比。体裁的学科分布不均降低了学科之间的可比性。尽管通用语料库不是以体裁代表性为建设标准，但是其学术文本部分存在的另外两个问题是语料常常不是完整的文章，而是文章片段，这不利于体裁结构分析等语篇研究，限制了应用的范围。可见，通用语料库未必适用于学术英语研究。正如Lee (2001: 37) 的观点，语境具体的专门语料库对于特殊用途英语研究具有更大的价值。专门语料库能够实现语料文本之间的可比性，便于基于体裁结构的文本考察，可以

在纯语言分析中加入互动、语用和语境因素。

第三，口语语料库和笔语语料库。

语料库按照语言传播媒介又可以分为口语语料库和笔语语料库。相对于笔语语料库，口语语料库建设难度较大，因为它受口语发生的即时多变环境影响较大，同时取样过程复杂困难。此外，口语语料库的建设还需要转写——把口语语料人工手动转写成文本，以文字形式录入语料库。为了满足各种研究需要，在转写口语语料时，通常还需要制定转写规则，规范记录口语特征(如停顿、犹豫、重复、强调等)的操作。因此，通常口语语料库学术英语研究在规模和数量上比笔语语料库学术英语研究相对要少。在学术英语研究方面，英国学术口语语料库(BASE corpus)和密歇根学术英语口语语料库(Michigan Corpus of Academic Spoken English, 简称MICASE)是两个具有权威和代表性的口语语料库，均以学生的学术英语口语为对象。以MICASE为例，该语料库包括学业辅导(advising session)、研讨会(colloquium)、论文答辩(dissertation defense)、小组讨论(discussion section)、访谈(interview)、实验交流(lab section)、课堂教学(lecture)、会议(meeting)、办公室交谈(office hour)、讲座(seminar)和论文报告(paper presentation)等，也囊括了常见的学术英语口语交流场景。与英国学术口语语料库(BASE corpus)和密歇根学术英语口语语料库(MICASE)相对应，英国学术笔语语料库(BAWE corpus)和密歇根大学高水平学生论文语料库(MICUSP)是较具代表性的笔语语料库，涵盖了学术英语语境中常见的笔语交流体裁。以MICUSP为例，该语料库包括论说文(argumentative essay)、创意写作(creative writing)、评论(critique/evaluation)、研究计划(proposal)、报告(report)、研究论文(research paper)等，涵盖学术英语书面体裁广泛的语言特征。

第四，本族语者语料库和二语学习者语料库。

学术英语语料库另一个常见的分类是本族语者语料库(native speaker

corpus) 和二语学习者语料库 (learner corpus)。本族语者语料库通常收集的是英语母语人群产出的学术英语语料, 而二语学习者语料库收集的则是英语作为第二语言的语料。在研究中, 我们常常把二语学习者语料库作为观察语料库, 以本族语者语料库为参照语料库, 进行对比分析, 目的在于发现学习者语言的特点。Sylviane Granger 提出的中介语对比分析方法 (Contrastive Interlanguage Analysis, 简称 CIA) 是较为常见的学习者语料库研究方法。

第五, 专家语料库和新手语料库。

专家语料库 (expert corpus) 和新手语料库 (novice corpus) 是学术英语研究方面独特且十分常见的语料库分类。专家语料库收集的是被认定为学术话语实践资深人群的语料, 代表学术英语话语共同体认可的语言特征, 而新手语料库收集的是在学术话语实践方面处于新手学徒阶段人群的语料, 代表有待被学术话语共同体接受的语言表现。学术英语与其他变体的英语语言不同, 它被学界认为不是任何人的母语 (Bourdieu & Passeron 1994; Hyland 2015 a)。因此, 正如 Swales (2004: 56) 所言, 学术英语语言使用最重要的区别不在于本族语和非本族语者, 而在于经验丰富、资深的学者与经验不足的初学者。Hyland (2005 b) 自建多学科本科毕业报告语料库和相应学科的权威期刊论文语料库。他将两个语料库分别视为新手语料库和专家语料库, 对比学生和学科专家在学术语篇中使用“读者带入” (reader engagement) 语言资源的异同。专家语料库和新手语料库有别于本族语者语料库和二语学习者语料库, 专家语料库未必全是本族语者语料, 而新手语料库也未必都是二语学习者语料。专家语料库和新手语料库都可以是本族语者语料。总之, 学者需要根据不同的研究目标来选择适宜类型的语料库。

第六, 共时语料库和历时语料库。

按照语料文本产生的时间, 语料库又可以分为共时语料库 (synchronic corpus) 和历时语料库 (diachronic corpus)。共时语料库收集的是同一时

期内产生的语言，而历时语料库收集的是不同时期内产生的语言。历时语料库建设的目的是为了研究语言的历时变化，比如西班牙拉科鲁尼亚大学建立的哲学英语文本历时语料库 (Corpus of English Philosophy Texts, 简称CEPhiT)，相关研究旨在分析中古近代哲学学术文本的语言特征 (Moskowich *et al.* 2016) 或者以共时语料库为参照发掘语言发展和变化的规律 (Seoane 2015)。就技术层面而言，相对于共时语料库，历时语料库的建设难度更大，面临的困难包括历史文本的搜集、历史文字的识别和转换等。

1.6.2 学术英语语料库建设

语料库的建设常常涉及诸多问题，如采集文本，取样和整理，考虑语料的代表性和均衡性，标注元信息等 (Sinclair 2005; 梁茂成等 2010)。对于学术英语语料库的建设而言，语料的代表性、均衡性以及元信息标注显得尤为突出。学术英语语料库研究讲求文本语境的专门性和具体性，比如文本所属学科越具体越好，文本体裁越明确越好，因此所建学术英语语料库需要能够充分代表聚焦的文本体裁，包含均衡的作者组成。

Leech (1991) 认为，语料是否具有代表性涉及研究结果能否普及到所关注的语言或语境的整体。McEnery *et al.* (2006) 认为，语料的代表性主要体现在文本的体裁范围和选取标准上。如果研究目标是学术英语的语言特征，那么语料在口笔语的体裁范围和构成上都需要均衡。就笔语语料而言，语料库不仅要包含研究论文、实验报告，还要包括教材、书评和学位论文等。如果研究者关注的是历史学期刊论文的语言特征，那么在建设语料库时需要考虑：是仅搜集研究论文，还是要包含书评和综述类文本。可见，语料的体裁范围取决于研究目的。此外，在建立期刊论文语料库时，我们还要全面地斟酌文本的选取标准。通常，一方面我们根据学科内期刊影响因子排名，如根据 ISI Web of Knowledge 发布的期刊影响因子参数，选定每个学科的期刊；另一方面，我们仍需要咨询学科专家，对排名上榜

的期刊进行评价并给予推荐建议。经由以上两步确认的期刊名单应该才是权威的、有说服力的。

学术文本的作者构成往往比较复杂。比如，期刊论文既有独立作者的文章，又包含多人作者的合作文章；既有本族语者的文章，又有非本族语者的文章。由于作者构成复杂，在语料库建设时，研究者需要思考如何准确地定位一种期刊。比如，对比中外学术期刊语言特征的研究经常将 *Chinese Journal of Applied Linguistics* 界定为国内刊物，但是该刊物采用的是国际同行评审，而且发表大量国际学者的学术论文。因此，在建设学术英语语料库时，我们不可忽视文本的作者构成。如果研究目标是学术论文中自称语的使用，那么作者构成要保持均衡，或者全部是独立作者，或者全部是多人作者，或者各占一半。