

语料库下的学术语篇话语分析——应用与挑战

曾艳

(湘潭大学 湖南湘潭 411105)

摘要:在语言学研究的广阔领域中,学术语篇的话语分析一直占据着重要的地位,这是因为学术语篇不仅承载着学科知识的传播与创新,还能体现作者与读者之间的交际互动。近年来,随着语料库语言学的发展,越来越多的研究者开始利用语料库技术对学术语篇进行深入分析,并揭示学术语篇中丰富的语言现象和交际特征,所以本文将探讨将语料库技术应用于学术语篇话语分析的前景和挑战,希望能够为学术语篇研究提供一些启示。

关键词:语料库;学术语篇;话语分析

引言:

语料库语言学作为现代语言学研究中的一项重要方法,其如今已然被广泛应用于各类语言现象的探讨中,尤其是在学术语篇的分析领域。应用语言学研究的核心任务之一就是探讨语言的实际使用情况,这不仅要求研究者能够通过传统的理论框架对语言现象进行分析,更需要依赖大量真实的语言数据来揭示规律和趋势,但是由于传统的语言分析方法往往局限于个别文本的分析,缺乏对大量语言数据的系统性挖掘,而语料库语言学的出现正是为了解决这一问题,通过构建大型语料库,并利用计算机技术对文本进行高效的分析,从而揭示语言使用中的共性特征。

一、语料库语言学与话语分析的关系

(一) 语料库语言学的基本概念

语料库语言学作为一种基于实际语言数据的研究方法,其主要是通过大量语言数据的统计分析去揭示语言使用的规律。期间语料库不仅可以提供语言研究的原始数据,而且还能通过计算机技术的辅助,使得对这些数据的分析更加精确和高效。与传统语言学方法相比,语料库语言学具有明显的优势:它避免了语言学家主观分析的偏差,能够从大量数据中发现语言现象的普遍性和特殊性,提供更加客观和全面的语言描述^[1]。而在语料库语言学中,学术语篇的分析则尤为重要,因为学术写作通常具有高度的规范性和系统性,语言使用上也有特定的结构要求。通过语料库分析,可以帮助研究者识别学术写作中的常见模式,并探讨不同学科、不同语域中的语言差异。

(二) 话语分析的基本理论与方法

话语分析是一种研究语言使用中的结构和意义的方法,其主要引用于语言在具体情境中的实际运作。其中话语分析不仅会涉及语言的形式结构,还包括其背后的社会、文化和认知因素。对于学术语篇的分析,话语分析注重对语言的衔接性、逻辑性、修辞手段等方面的解读,而且话语分析不仅需要理解个别的语句,还需要分析不同语言成分之间的关系和整体语篇的构建,近年来随着计算机技术的发展,语料库语言学已然逐渐成为话语分析的重要工具,研究者们通过语料库的方法可以高效地处理大量文本数据,揭示学术语篇中的共性特征,从而为学术写作的规范化提供实证支持^[2]。

二、语料库技术在学术语篇话语分析中的应用

(一) 语料库建设与数据预处理

在学术语篇话语分析中,研究者往往需要根据研究目的和研究对象,去选择合适的语料来源,如学术期刊、会议论文、学位论文等。在收集到足够的语料后,还需要进行一系列的数据预处理工作,如文本清洗、分词、词性标注、句法分析等,以确保数据的准确性和可用性。例如,在构建应用语言学学术语篇语料库时,研究者可以选取国内外知名应用语言学期刊上发表的论文作为语料来源,并通过自动化工具和人工校对相结合的方式,对语料进行清洗和标注,确保数据的准确性和一致性;与此同时还可以根据研究需要,对语料进行进一步的分类

和标注,如按学科领域、研究主题、语体风格等进行分类,以便后续的分析和比较。

(二) 话语标记分析

话语标记是学术语篇中常见的语言现象,它们在构建语篇连贯性、表达作者态度等方面发挥着重要作用,研究者利用语料库技术可以系统地分析学术语篇中的话语标记,揭示其使用规律和交际功能。其中在话语标记分析中,研究者可以先确定研究范围和分析框架。例如,可以选择表示转折、因果、条件等关系的话语标记作为研究对象,分析其在学术语篇中的出现频率、搭配关系、语义功能等,然后通过自建语料库和统计软件的支持去方便地提取出这些话语标记的实例,并进行量化分析和质性描述^[3]。例如,有研究发现,在学术语篇中,“但是”“然而”等转折话语标记的使用频率较高,且常常与表示对比、反驳等语义关系的句子搭配使用。这些话语标记不仅有助于构建语篇的连贯性,还能够表达作者对前人研究或不同观点的态度和立场,由此可见通过深入分析这些话语标记的使用情况,可以进一步揭示学术语篇中作者与读者之间的交际互动。

(三) 语步结构分析

语步结构作为学术语篇的重要组成部分,它可以直接反映出作者如何组织和呈现自己的研究成果和观点,研究者利用语料库技术可以对学术语篇的语步结构进行量化分析和质性描述,揭示其共性和差异。所以在语步结构分析中,研究者可以先确定研究范围和分析框架。如,可以选择学术论文的引言、方法、结果、讨论等部分作为研究对象,分析它们在学术语篇中的出现顺序、篇幅比例、语言特征等,这样通过自建语料库和统计软件的支持,研究者们变可以系统地提取出这些部分的实例,并进行量化分析和质性描述^[4]。例如,在学术语篇中,引言部分通常用于介绍研究背景、目的和意义,方法部分用于描述研究设计和实验过程,结果部分用于呈现研究数据和发现,讨论部分用于解释结果和提出观点。这些语步结构不仅能够帮助作者更清晰地表达自己的研究成果和观点,还能够引导读者理解和评价这些成果和观点。

三、语料库技术在学术语篇话语分析中的挑战

(一) 语料库建设的挑战

语料库技术在学术语篇话语分析中的首要挑战在于语料库的建设,毕竟语料库作为数据分析的基础,其质量和代表性直接影响分析结果的准确性和有效性。然而,在构建高质量的学术语篇语料库时,研究者常常面临多重困难;其次学术语篇的多样性和复杂性会增加语料库建设的难度,且不同学科、不同领域的学术语篇在语言风格、结构布局、词汇使用等方面存在明显差异。因此,在构建语料库时,研究者需要充分考虑这些因素,确保语料的广泛性和代表性,而要想做到这一点,那么就要求研究者不仅要具备深厚的学科知识背景,还要具备跨学科的视野和能力,以全面收集和整理学术语篇语料^[5]。另外,语料的获取和标注也是语料库建设中的一大挑战,学术语篇通常分散在不同的学术期刊、会议论文、学位论文等来源中,获取

这些语料需要耗费大量时间和精力；与此同时语料的标注也是一项繁琐且复杂的工作，需要研究者具备专业的语言学知识和标注技能，因为标注的准确性会直接影响后续数据分析的结果，所以如何保证标注的一致性和可靠性成为语料库建设中的一大难题。

（二）数据处理的挑战

在构建好语料库之后，研究者还需面对数据处理的挑战，因为学术语篇话语中会分析涉及大量的语言数据，所以如何高效地处理这些数据并从中提取有用的信息成为研究者必须解决的问题。其中数据处理的挑战主要体现在数据的清洗和预处理上，毕竟学术语篇中常常包含各种噪声数据，如无关的注释、错误的拼写、不规范的标点符号等，这些数据会影响后续数据分析的准确性。因此，在数据处理阶段，研究者需要对语料进行仔细的清洗和预处理，以消除这些噪声数据的影响。然而，这过程往往会耗时费力，且需要研究者具备丰富的语言学知识和经验。除此以外，数据处理还面临数据规模的问题，因为如今随着学术语篇数量的不断增加，语料库的规模也随之不断扩大，所以如何高效地处理大规模数据并从中提取有用的信息成为了所有研究者必须面对的挑战。可是传统的数据处理方法可能无法适应大规模数据的需求，因此研究者需要探索新的数据处理技术和方法，以提高数据处理的效率和准确性^[6]。

（三）分析方法的挑战

不同的分析方法会适用于不同的研究目的和数据类型，所以如何选择合适的分析方法成为研究者必须考虑的问题，其中学术语篇话语分析涉及多个层面和维度，如词汇使用、句法结构、篇章布局、交际意图等。不同的研究方法可能适用于不同的分析层面和维度。例如，词汇共选分析适用于揭示词汇之间的关联和搭配关系；语步结构分析适用于揭示学术语篇的组织结构和逻辑顺序；话语标记分析适用于揭示学术语篇中的交际意图和语用功能，但是这些方法各有优缺点，如何根据研究目的和数据类型选择合适的方法成为研究者必须面对的挑战。而且分析方法的挑战还体现在跨学科的融合上，因为学术语篇话语分析不仅会涉及语言学知识，还涉及认知科学、社会学、心理学等多个学科领域。所以如何将不同学科领域的研究方法和技术相结合，以全面揭示学术语篇的深层结构和交际意图成为是每位研究者必须要探索清楚的问题。

四、语料库技术在学术语篇话语分析中的复杂性及发展前景

（一）话语分析的复杂性

学术语篇话语分析作为一项复杂而多维的任务，其复杂性主要体现在以下几个方面。

1.学术语篇的语言特征具有独特性：学术语篇通常会使用专业术语、复杂句式和严谨的逻辑结构，这会使得其语言分析难度大大增加。例如，在学术语篇中，作者常常使用隐喻、借代等修辞手法来增强论述的说服力，这些修辞手法的识别和分析需要研究者具备深厚的语言学功底和敏锐的学术洞察力。

2.学术语篇的交际意图和语境背景是话语分析的重要方面：学术语篇不仅仅是作者个人思想的表达，其更是作者与读者之间的一种交际互动，因为在学术语篇中，作者常常会通过特定的语言策略来构建自己的学术立场和观点，同时考虑读者的认知背景和学术需求，因此在话语分析过程中，研究者需要充分考虑学术语篇的交际意图和语境背景，以揭示其深层的交际意义和语用功能。

3.学术语篇的话语结构也会增加分析的复杂性：学术语篇通常由多个部分组成，如引言、方法、结果、讨论等，每个部分都有其特定的功能和语言特征，所以在话语分析过程中，研究者需要识别和理解这些部分之间的逻辑关系和语言衔接，以

揭示学术语篇的整体结构和组织模式。

（二）语料库技术的应用前景

尽管语料库技术在学术语篇话语分析中面临诸多挑战，但其应用前景依然广阔，毕竟如今随着计算机技术的不断发展，语料库技术的自动化和智能化水平将不断提高。例如，自然语言处理（NLP）技术的进步将使得语料库的数据收集、标注和预处理工作更加高效和准确。机器学习（ML）和深度学习（DL）算法的应用也将为话语分析提供更强大的工具和方法，使研究者能够更深入地挖掘学术语篇中的语言规律和交际特征；除此以外语料库技术在跨学科研究中的应用前景也值得关注，因为学术语篇话语分析不仅涉及语言学知识，还涉及认知科学、社会学、心理学等多个学科领域。而语料库技术则可以为这些学科领域提供丰富的语言数据和分析工具，促进不同学科之间的交叉融合和创新发展。未来相信随着语料库技术的不断完善和普及，其在学术语篇话语分析中的应用将更加广泛和深入，语料库将成为学术语篇研究的重要资源和工具，为揭示学术语篇的语言规律和交际特征、推动学术交流和知识创新发挥重要作用；与此同时我们也期待看到更多基于语料库技术的创新研究方法和成果涌现出来，为学术语篇话语分析带来新的突破和发展。

结语：

总而言之，语料库技术在学术语篇话语分析中的应用展现出了其强大的潜力和广阔的前景，尽管其如今依旧面临着语料库建设、数据处理和分析方法等多方面的挑战，但随着计算机技术的不断进步和跨学科研究的深入发展，这些挑战将会被逐渐克服。从而不仅能够为学术语篇研究提供新的视角和工具，还可以促进语言学与其他学科之间的交叉融合。未来，语料库技术将继续在学术语篇话语分析中发挥重要作用，推动学术交流和知识创新的深入发展，为语言学研究 and 学术写作提供更加全面和深入的支持。

参考文献：

- [1]余强.基于 VOSviewer 的国内语料库语言学研究(2004-2024)[J].语言与文化研究,2024,32(06):6-10.
 - [2]王倩.框架语义学视域下应用语言学国际期刊学术论文中“兴趣”标记的历时考察[J].中国 ESP 研究,2024,(03):93-105+177.
 - [3]凡颜.基于语料库的应用语言学学术写作摘要的短语复杂度研究[J].语言与文化研究,2024,32(05):23-27.DOI:10.19954/j.cnki.lcr.2024.05.057.
 - [4][1]周茜.语料库语言学在话语分析研究中的作用探讨[J].课程教育研究,2018(40):2.DOI:CNKI:SUN:KCJY.0.2018-40-010.
 - [5]周欣.基于语料库的应用语言学和医学学术英语词块比较研究[J].海外英语,2024,(05):67-69.
 - [6]唐丽萍.语料库语言学在批评话语分析中的作为空间[J].外国语,2011,34(4):43-49.
 - [7]高蓓.批评话语分析中的语料库语言学方法研究[J].中国多媒体与网络教学学报旬刊,2018(04):36+51.DOI:CNKI:SUN:JMNT.0.2018-04-017.
 - [8]郭敏.基于语料库语言学的批评话语分析[J].英语广场:学术研究,2017(5):2.
- 曾艳(1981-),女,汉族,硕士研究生,讲师,研究方向为社会语言学,学术英语。
- 湖南省教育厅一般项目《基于语料库的美国媒体涉华疫情报道的批评话语分析》(22C0061);湖南省教育科学规划课题《学科教学知识视角下学术英语教师教学学术能力发展研究》(XJK22CGD009)