

基于语料库语言学的学术文本摘要体裁分析研究

安亚楠, 李晶洁

东华大学, 上海

收稿日期: 2022年10月1日; 录用日期: 2022年10月26日; 发布日期: 2022年11月4日

摘要

语料库语言学自20世纪60年代以来深受行为主义的影响, 至今已有近六十余年的发展。近年来, 语言学研究更加注重跨学科的研究, 而语料库语言学与体裁分析之间的交叉研究也是重要研究方向之一。本文收集了30篇关于语料库语言学的学术文本(RA)摘要, 并将其根据两种不同的体裁分析模型进行研究, 即IMRD和CARS模型, 从文本化、结构化和语境化三个角度切入语步分析。文章目的在于了解关于语料库语言学的学术文本摘要的通用结构潜力, 本文希望通过体裁分析理论在语料库语言学的学术文本摘要中的实际应用, 促进语料库语言学和体裁分析两个研究方向之间的紧密联系。

关键词

体裁分析, 语料库语言学, 学术文本, 摘要

A Genre Analysis of Research Article Abstract on Corpus Linguistics

Ya'nan An, Jingjie Li

Donghua University, Shanghai

Received: Oct. 1st, 2022; accepted: Oct. 26th, 2022; published: Nov. 4th, 2022

Abstract

Influenced by behaviorism, corpus linguistics has been developing for nearly 60 years since 1960s. In recent years, Linguistic studies pay more attention to the interdisciplinary research and the research between corpus linguistics and genre analysis is no expectations. This paper collects thirty research article (RA) abstracts on corpus linguistics and classified them into two different models

of genre analysis, namely IMRD and CARS model put forwarded by Swales, and then focuses on the move analysis from different perspectives, that is from the aspects of textualization, organization and contextualization of the text. Furthermore, the aim of this paper is to know about the generic structural potential of the RA abstracts on corpus linguistics. Finally, this paper hopes to promote a closer connection between these two linguistic research directions through the practical application of the theory of genre analysis in analyzing the RA abstracts on corpus linguistics.

Keywords

Genre Analysis, Corpus Linguistics, Research Article (RA), Abstracts

Copyright © 2022 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

随着体裁分析研究的不断发展, 体裁分析不仅涉及文体分析, 还涉及话语分析, 同时它也是跨学科研究的产物, 其根本目的是为了学习如何使用语言的策略及交际[1]。随着体裁分析逐渐成为语言学的热门研究内容, 其应用研究方向主要包括以下几点。从国外研究领域来看, 一方面, 国外学者主要关注体裁教学法在 ESP/EAP 教学中的实践研究, 教师在教学中逐渐将社会情境考虑在内并深刻了解到体裁教学理论的重要性。教师认为体裁分析是以交际为目的的, 因此体裁分析教学可以使学生更好地理解什么是体裁, 而这些交际活动促进了教学中对专业文本的理解[2]。同时, 国外学者将目光放在学术体裁的相关研究, 通过体裁分析的研究过程, 发现期刊类别和硕士论文类别的摘要中使用语步频率是不同的, 除了摘要部分, 国外学者也对学术论文的其他部分进行了体裁分析研究, 研究发现 Swales 提出的 IMRD 模型的多样使用结果, 每个部分都有自己独特的组织结构。最后, 近来随着互联网技术的发展, 他们更加关注网络平台上的不同体裁, 如百科全书网站、社交媒体或各大学网站, 研究文本类型丰富, 成果颇丰。与国外学术界体裁分析的研究进展相比, 国内体裁分析发展相对缓慢。国内学者对体裁分析的研究主要集中在两个方向, 一个是基于体裁分析理论的体裁应用研究, 另一个是政治材料的体裁分析。针对第一个研究方向, 国内学者非常重视网络平台、政治材料和学术论文的体裁分析, 尤其是在学术论文方面; 对于第二个研究方向, 国防话语的研究最近成为一个新兴热门领域。例如, 国内学者从 CARS 模型出发, 提炼出韩国国防白皮书前言的结构过程, 构建了四语步、十四层次的新型结构框架, 也有学者从话语分析的角度关注南海仲裁案[3]。并分析了南海仲裁案的体裁互文性, 认为该仲裁案从一开始就突破了“仲裁”体裁的界限, 大规模“挪用”了“诉讼”的体裁资源, 以达到妖魔化中国、损害中国南海主权的企图。从以上研究内容中可以看出, 体裁分析的发展涉及多个领域并趋于成熟[4]。但同时也有需要继续改进的地方, 例如国内大部分研究主要集中在语步分析上, 研究的深度相对不足, 即对语言层面的功能和体裁特征没有进一步研究。

2. 文献综述

2.1. 体裁分析与语料库语言学

体裁(genre)是“一种交际事件, 其参与者持有某种相同的交际目的”(a class of communicative events, the members of which share some set of communicative purposes) [5]。体裁分析(genre analysis)描述话语社团

成员如何“将特定的体裁标记与特定的交际事件相对应”(assign a particular genre label to a particular communicative event) [6]。体裁(genre)一词源于希腊语中的 category 以及 classification, 如今它的指称意义在很大程度上已经扩大, 进而指的是人类社会交流活动中几乎所有的言语事件。而体裁分析是对特定体裁语篇的宏观结构和交际功能进行分析。它能够把篇章的语言表达形式与其交际目的及使用的社会场合结合起来, 通过对其深层交际目的的揭示来描述与解释某一特定体裁的结构和语言特征。

语料库语言学是 20 世纪 80 年代以来逐渐发展起来的一门学科, 它主要用于文本数据的检索、取样、分析和统计。语料库语言学涵盖了第二语言习得、外语教学与研究、翻译服务与实践、话语分析、语言处理等多个语言学的专业性和综合性领域, 随着语料库语言学的兴起, 短语学已成为语言学的一个专门学科领域。它聚焦于词语的组合行为, 尤其关注词语的形式、意义和功能的共选行为特征。共选理论的基本观点是: 在语言交际中, 形式与意义是共选的, 二者密不可分。简言之, 交际过程中最基本的共选关系有两个: 词汇和语法的共选、词汇和词汇的共选。前者认为, 为表达给定意义, 一旦具体的词形被选定, 其语法结构也随之选定; 后者说的是, 一个词形的选定直接影响到下一个词形的选择, 从而导致多个词一同被选。这两种基本的共选关系又会导致另外一些更高层面的共选关系, 如由词汇和语法相结合而成的型式与意义的共选, 等等。共选理论尤其强调词汇和语法不可分离的紧密关系, 强调词语的线性共选行为对表达意义的作用[7]。以 John Sinclair 为代表的语言学家基于语料库证据探讨了这些共选行为, 提出了短语意义单位: 意义的最重要承载单位是短语, 而不是单个词。这一系列理论和体裁分析理论不谋而合, 他们都强调语境的重要性。语料库语言学不断吸引不同研究背景和方向的众多学者加入语料库语言学的研究队伍, 使其得到了空前的发展, 体裁分析作为语言学研究的热点话题, 近年来与语料库语言学的关系更为密切。大多数研究者倾向于将语料库语言学的方法与体裁分析的理论相结合, 如语步分析理论等。

2.2. 学术文本摘要研究

摘要是对整篇论文的缩写和准确表述, 它是作者想要表达的主要观点。摘要是论文的总揽, Hyland 曾经指出, 一个好的摘要可以让论文的潜在读者相信作者有讨论这个领域问题的专业能力, 从而吸引读者阅读整篇文章。在最近的研究热点中, 摘要的相关研究越来越重要且普遍, 如“科技论文英文摘要中词块的语步特征”这篇文章, 作者将科技论文摘要作为体裁分析文本进行语步分析, 是体裁分析的重要研究方向之一[8]。国内外学者围绕不同学科论文摘要的研究证实, 不同学科中存在体裁差异, 不同体裁中反映学科差异, 而这些差异是学科性质和特点所导致的。摘要的写作风格和语篇结构, 甚至语言特征受到了学科文化和特征的深刻影响[9]。

2.3. 语步分析——CARS 模型及 IMRD 模型

体裁分析中的语步概念是指话语或修辞单位, 它在书面或口语话语中发挥着具有一定连贯性的交际功能。尽管它有时与语法单位如句子、语篇或段落相一致, 但就其语言实现功能而言, 它是相对灵活的。一方面, 它可以由有且只有一个句子实现; 另一方面, 它可以由多个句子实现功能。它是一个功能单位, 而不是一个形式单位[10]。国外研究学者和方向主要分为两个不同的体裁分析学派。以 Swales 和 Bhatia 为代表的斯韦尔斯学派和以 Martin 为代表的澳大利亚学派, 都隶属于以 Halliday 为代表的系统功能语言学的分支, 即体裁分析理论研究。

语步分析模型主要包括 Swales 提出的两种不同类型的模型, 即 IMRD 模型和 CARS 模型。IMRD 模型包括四语步。首先, I (Introduction)指引言, 主要介绍作者的研究目的、论点或假设, 也可能包括作者想要解决的问题; M (Methodology)指方法论, 讲述作者在论文中使用什么方法和研究过程; R (Results)

指结果,这是摘要中非常重要的部分,作者将在结果部分提供最新研究结果;D(Discussion)指讨论,主要集中在展示结果后的讨论。而CARS模型相比IMRD模型来说,相对复杂,CARS模型是Create a Research Space的缩写,它由语步及不同步骤组合组成。第一步S1明确领域,即文章关注的研究领域,之后建立利基,S1A是指出目前研究缺口,或使用S1B对已知的内容进行补充说明;第二步S2阐述此项研究的必要性及重要性。第三步S3要求笔者通过S1来介绍目前的论文进展工作,即以描述性或目的性的方式明确目前的研究进展,即S2提出相关问题或假设,下一步S3是对文章研究内容进行定义说明,S4总结方法,S5是宣布主要成果,这一步在摘要中是相对必要的,以显示作者的最终论文成果,摘要最后一部分S6则要求作者说明目前研究的价值。

3. 研究方法

本文主要分为两个步骤。第一步是样本收集,第二步是语步识别和分析。笔者根据三条基本规则进行样本收集,即同类性、随机性和详尽性。该样本属于同一学术领域,即语料库语言学学术论文。笔者在收集样本时,将研究领域确定为语料库语言学,并将语料库语言学的两本权威期刊作为语料来源,即*Corpus Linguistics and Linguistic Theory*和*International Journal of Corpus Linguistics*。笔者从*International Journal of Corpus Linguistics*中选取了15篇2020年发表的RA摘要,从*Corpus Linguistics and Linguistic Theory*中选取了7篇2020年发表的RA摘要,由于2020年发表的论文不够多,笔者又增加了8篇2019年发表的RA摘要作为样本进行了分析。第二步,语步识别和分析,笔者首先根据IMRD模型和CARS模型对上述30个文本样本进行标记。在这个过程中,笔者区分了M1和M3S1的区别,它们在某种程度上功能是完全一致的,之后分别完成了30个样本的语步识别。其次,笔者综合了上述数据并制作了相关的图表,最后通过不同的角度进行分析,即从结构角度(宏观方面)说明RA摘要是属于IMRD模型还是CARS模型。而从文本化角度(微观方面),其中包括时态和人称代词的分析,最后分析通用结构的潜力[11]。

4. 结果及讨论

本文通过对30个样本中IMRD和CARS模式的使用频率分析,以了解语料库语言学的学术文本摘要更倾向于哪种模式。本章主要从两个部分进行阐述,即学术文本摘要的结构及文本化,同时也是从宏观和微观角度进行分析。

4.1. 学术文本摘要的结构研究(宏观角度)

4.1.1. IMRD及CARS模型分类

经过语步识别,RA的摘要主要分为IMRD模型还是CARS模型。在30个样本中,IMRD模型的频率为14,CARS模型为16,这表明IMRD的数量只占40%,而CARS模型的数量占60%。从这样的分析结果中,可以得出语料库语言学的学术论文摘要更倾向于使用CARS模型,这也在一定程度上显示了通用结构的潜力。如下图表1。

Table 1. The proportion of IMRD and CARS model data
表 1. IMRD 及 CARS 模型占比

Model	IMRD	Cars Model
amount	14	16
percentage	40%	60%

4.1.2. CARS 模型分析

下图表格显示了 CARS 模型中不同语步的频率。由此可知 M3S4 在关于语料库语言学的学术文本摘要中使用频率最高, 这说明几乎每篇摘要都包含研究方法, 方法论是摘要中不可或缺的一部分。排名第二的是 M3S1, 即以描述性或目的性的方式宣布研究成果, 由此可见论文的描述和目的讨论在学术文本摘要中仍然是重要的。图表中排名第三的是 M1, 即在撰写论文前确定自己的领域。根据研究数据, 其重要性顺序排序如下图表 2。

M3S4 > M3S1 > M1 > M3S5 > M2S1A/M3S6 > M2S1B > M3S3 > M3S2/M3S7/M3S8

Table 2. The distribution of moves

表 2. 语步分布

Moves	Amount	Percentage
M1	14/18	78%
M2S1B	6/18	33%
M2S2	0	0%
M3S1	15/18	83%
M3S2	1/18	5%
M3S3	3/18	17%
M3S4	16/18	89%
M3S5	8/18	44%
M3S6	7/18	39%
M3S7	1/18	5%
M3S8	1/18	5%

4.2. 学术文本摘要的文本化研究(微观角度)

4.2.1. 时态

通过分析数据, 学术文本主要存在三种类型的时态, 即一般现在时、一般过去时和现在完成时。根据下面列出的数据, 可以了解到 CARS 模型的每一个语步都有一般现在时存在, 频率最高, 占比 100%, 之后第二频率出现的是现在完成时, 最后是一般过去时。

M1:	Simple Present (10/14)/Present Perfect (3/14)/Simple Past (1/14)
M2S1A:	Simple Present (3/6)/Simple Past (2/6)/Present Perfect (1/6)
M2S1B:	Simple Present (3/6)/Simple Past (2/6)/Present Perfect (1/6)
M3S1:	Simple Present (14/14)
M3S2:	Simple Present (1/1)
M3S3:	Simple Present (3/3)
M3S4:	Simple Present (14/15)/Simple Past (1/15)
M3S5:	Simple Present (7/8)/Simple Past (1/8)
M3S6:	Simple Present (8/9)/Simple Future (1/9)
M3S7:	Simple Present (2/2)
M3S8:	Simple Present (1/1)

4.2.2. 人称代词

许多学术英语研究表明, 人称介入是一种人际交往策略, 它服务于学术文本的交际目的, 即作者作为研究者, 如何协调与同行业的读者和所属学科的关系[10]。根据相关研究背景, 我们应该高度重视学术文本摘要中人称代词的研究。根据详细的数据可见除了语步 M3S2/M3S4/M3S8 之外, 作者在撰写文章主题及目的时倾向于使用第三人称代词, 但当作者在介绍文章的研究方法时(M3S4), 第一人称或第三人称代词的使用频率是相同的。

M1:	Third [topic] (14/14)
M2S1A:	Third [topic] (7/7)
M2S1B:	Third [topic] (6/6)
M3S1:	Third [topic/paper/purpose] (11/15)/First [I/we] (4/15)
M3S2:	First [we] (1/1)
M3S3:	Third [topic] (3/3)
M3S4:	First [I/we] (8/16)/Third [study/samples] (8/16)
M3S5:	Third [study] (7/8)/First [We] (1/8)
M3S6:	Third [study] (9/9)
M3S7:	Third [study] (1/1)
M3S8:	First [I] (1/1)

5. 结论

本文重点关注体裁分析与语料库语言学的交叉研究, 通过分析 30 个样本中 IMRD 和 CARS 模型的使用频率, 来了解语料库语言学的学术论文摘要更倾向于哪种模型, 同时也能更好地了解作者如何利用不同的语法结构来实现体裁分析的目的, 即交际目的。通过对所有相关数据的分析, 我们可以得到这样的结论: 语料库语言学学术论文摘要的作者倾向于使用 CARS 模式, 其中 M3S4 的频率最高, 其次是 M3S1 和 M1, 这说明在每个学术论文摘要中, 作者应该在一定程度上包含论文的描述及目的。而在微观方面, 语料库语言学的学术论文摘要更倾向于使用第三人称代词和一般现在时, 以显示其研究的客观性, 同时也关注并了解到更多关于语料库语言学学术论文摘要的通用结构潜力。

参考文献

- [1] de Saussure, F. (1971) *Cours de Linguistique Générale*. Payot, Paris.
- [2] Hyland, K. (2000) *Disciplinary Discourses: Social Interactions in Academic Writing*. Longman, London.
- [3] Ren, H.W. and Li, Y.Y. (2011) A Comparison Study on the Rhetorical Moves of Abstracts Published Research Articles and Master's Foreign-Language Theses. *English Language Teaching*, 4, 162-166. <https://doi.org/10.5539/elt.v4n1p162>
- [4] 郭苗. 基于语料库的中外应用语言学期刊摘要体裁的对比分析[D]: [硕士学位论文]. 西安: 陕西师范大学, 2019.
- [5] Swales, J.M. (1990) *Genre Analysis: English in Academic and Research Settings*. Cambridge University Press, Cambridge.
- [6] Paltridge, B. (1997) *Genre, Frames and Writing in Research Settings*. John Benjamins Publishing Company, Philadelphia. <https://doi.org/10.1075/pbns.45>
- [7] 卫乃兴. 共选理论与语料库驱动的短语单位研究[J]. 解放军外国语学院学报, 2012, 35(1): 1-6.
- [8] 刘冰. 基于语料库的科技论文英文摘要中词块的语步特点研究[J]. 河南理工大学学报, 2020, 21(5): 59-66.
- [9] Sinclair, J.M.H. (2004) *Trust the Text: Language, Corpus and Discourse*. Routledge, London and New York.
- [10] 李晶洁, 周昕玥. 学术语篇“功能句干”的引导元话语特征分析[J]. 外语教学理论与实践, 2018(1): 64-73.

- [11] Tseng, F. (2011) Analyses of Move Structure and Verb Tense of Research Article Abstracts in Applied Linguistics. *International Journal of English Linguistics*, **1**, 27-39. <https://doi.org/10.5539/ijel.v1n2p27>