

湖北省社会科学基金项目成果

中国学者国际期刊论文 语言复杂度研究

A study on Linguistic Complexity of
International Journal Publications by Chinese Scholars

吴雪 著



WUHAN UNIVERSITY PRESS

武汉大学出版社

图书在版编目(CIP)数据

中国学者国际期刊论文语言复杂度研究/吴雪著. —武汉:武汉大学出版社, 2018. 10

ISBN 978-7-307-20529-1

I. 中… II. 吴… III. ①学术期刊—论文—语言—研究—中国
②学术期刊—研究—世界 IV. ①G312 ②G237.5

中国版本图书馆 CIP 数据核字(2018)第 210148 号

责任编辑:郭 静

责任校对:汪欣怡

版式设计:汪冰滢

出版发行: **武汉大学出版社** (430072 武昌 珞珈山)

(电子邮件: cbs22@whu.edu.cn 网址: www.wdp.com.cn)

印刷: 北京虎彩文化传播有限公司

开本: 720 × 1000 1/16 印张: 12 字数: 173 千字 插页: 1

版次: 2018 年 10 月第 1 版 2018 年 10 月第 1 次印刷

ISBN 978-7-307-20529-1 定价: 35.00 元

版权所有, 不得翻印; 凡购我社的图书, 如有质量问题, 请与当地图书销售部门联系调换。

前 言

由于语言多元化和言语复杂性，语言复杂度一直以来都是一个多学科、多维度的研究主题。语言学领域语言复杂度研究一直以来都是热点研究内容之一，虽然学者们针对语言复杂度展开的研究已斗量筭计，但由于研究设计、研究对象、研究方法等差异，相关问题至今仍有待探索。此外，先前研究大多仅限于语言复杂度某一方面的研究，并未考虑特定语篇语言规范下语言复杂度与语言产出质量的具体关联，以至于大部分外语教学和实践一味追求复杂单词、词组和句法的输入与输出。因此，本书试图采用语料库语言学的理论与方法对学术语篇语言复杂度与文本可读性展开研究。

本研究内容之一是探索中国学者国际期刊英语论文词汇复杂度特征及其对文本可读性的预测力。我们先分别在中国学者和本族语学者国际期刊英语论文语料库中开展词汇复杂度对文本可读性的预测力对比研究，并在此基础上建立词汇复杂度对文本可读性预测力模型。从对比结果来看，中国学者期刊论文文本可读性显著低于本族语学者，中国学者与本族语学者期刊论文中的词汇复杂度在某些维度指标上存在显著差异。比如，词汇密度方面，中国学者期刊论文词汇密度平均秩次显著低于英语本族语学者，即除了语法词汇以外，中国学者期刊论文中包含更少的动词、名词、形容词、副词等。从建立的模型来看，进入中国学者和本族语学者模型的词汇复杂度指标不尽相同，模型中各词汇复杂度指标系数也存在差异，从中国学者模型来看，对期刊论文文本可读性测量值 FRE 和 SMOG 具有最强预测力的词汇复杂度指标都是对数类符形符比，而从本族语模型来看，对期刊论文文本可读性值 FRE 和 SMOG 具

有最强预测力的词汇复杂度指标都是修饰语变化度。

本研究内容之二是探索中国学者国际期刊英语论文句法复杂度特征及其对文本可读性的预测力。同样地,我们又分别在中国学者和本族语学者国际期刊英语论文语料库中进行句法复杂度对文本可读性的预测力对比研究,也建立了相关预测力模型。从对比结果来看,中国学者与本族语学者期刊论文中句法复杂度在五个维度部分指标上存在显著差异。比如,在语言产出长度上,中国学者期刊论文的平均句子长度、平均T单位长度、平均子句长度三项指标值均显著高于本族语学者,即中国学者期刊论文语言产出长度显著高于本族语学者。从建立的模型来看,平均句子长度、T单位中动词短语数量、T单位中子句数量、T单位中并列短语数量在中国学者和本族语者期刊论文中均对文本可读性有较强预测力。此外,从中国学者模型中各句法复杂度指标系数来看,对期刊论文文本可读性FRE值具有最强预测力的句法复杂度指标是T单位中从属句数量,对文本可读性SMOG值最具有预测力的是子句中并列短语数量。

在前两个研究内容基础上,我们又引入学科变量,以进一步探索不同学科中国学者国际期刊英语论文词汇复杂度和句法复杂度特征及其对文本可读性的预测力,并对前两个研究内容中建立的模型进行补充完善。对比研究发现(1)理工学科期刊论文比人文社科可读性更强;(2)词汇复杂度方面,中国学者不同学科期刊论文的词汇复杂度各项指标和文本可读性差异主要存在于理工学科和人文社科之间:理工科论文中高级词汇数量更多,而文科专业动词词汇更丰富复杂;(3)句法复杂度方面,各学科在并列结构和复合名词结构上差异效应量较小或不具备显著差异,但是理工科专业期刊论文语言产出长度在平均句子长度和平均T单位长度两方面显著低于人文社科学科,此外,理工学科期刊论文中从属结构和动词结构也少于人文社科学科。根据以上对比结果,我们对已建立的模型中各变量系数进行了重新修正。

此外,在各研究内容定量统计分析和模型构建后,我们还结合话语分析理论和方法,从学术语篇视角对研究结果进行了定性分析,以期能

揭示语言特征背后的语言生成和使用机制。最后，我们结合研究结果对词汇复杂度、句法复杂度、文本可读性、学科等的关系与相互作用进行了综合分析，并从英语作为通用语视角进行了初步解释。

本书以国际期刊论文为主要研究对象，对学术语篇语言复杂度展开研究，尤其对词汇复杂度、句法复杂度、文本可读性之间的关系进行了较为深入的探索，并建立了相关预测力模型，以期为相关研究领域研究者提供理论和方法参考。在开展本书研究中，我的博士生导师华中科技大学雷蕾教授给予了最大的支持、指导和鼓励，雷蕾教授课题组成员们也都积极参与本研究数据收集和整理工作，在此向他们表示衷心的感谢。同时，也感谢我的家人一直以来的默默陪伴和源源不断的鼓励，愿你们安康。

目 录

第一章 绪论	1
1.1 研究背景	1
1.2 研究问题	5
1.3 研究意义	7
1.4 本书结构	8
第二章 文献综述	10
2.1 国际期刊论文写作研究	10
2.1.1 系统功能语言学视阈下的国际期刊论文写作研究	11
2.1.2 学术语言能力理论视阈下的国际期刊论文写作研究	12
2.2 词汇复杂度研究	15
2.2.1 词汇复杂度缘起	15
2.2.2 词汇复杂度测量方法	16
2.2.3 词汇复杂度研究综述	19
2.3 句法复杂度研究	21
2.3.1 句法复杂度缘起	21
2.3.2 句法复杂度测量方法	22
2.3.3 句法复杂度研究综述	24
2.4 文本可读性研究	26
2.4.1 文本可读性缘起	26
2.4.2 文本可读性测量方法	27
2.4.3 文本可读性研究综述	29

2.5 本章小结.....	32
第三章 概念框架和研究设计	36
3.1 重要概念定义及测量指标界定.....	36
3.1.1 词汇复杂度定义及测量指标	36
3.1.2 句法复杂度定义及测量指标	39
3.1.3 文本可读性定义及测量指标	40
3.2 研究设计.....	43
3.2.1 研究数据	43
3.2.2 数据分析工具	44
3.2.3 数据分析方法	46
3.3 本章小结.....	50
第四章 词汇复杂度各维度指标对文本可读性预测力	51
4.1 中国学者与英语本族语学者期刊论文词汇复杂度及文本可读性差异.....	51
4.2 中国学者期刊论文词汇复杂度对文本可读性预测力.....	58
4.2.1 中国学者期刊论文词汇复杂度与 FRE 回归分析.....	58
4.2.2 中国学者期刊论文词汇复杂度与 SMOG 回归分析	62
4.3 本族语学者期刊论文词汇复杂度对文本可读性预测力.....	65
4.3.1 本族语学者期刊论文词汇复杂度与 FRE 回归分析	65
4.3.2 本族语学者期刊论文词汇复杂度与文本可读性 SMOG	68
4.4 中国学者与本族语学者期刊论文词汇复杂度与文本可读性：文本实例.....	71
4.5 本章小结.....	80
第五章 句法复杂度各维度指标对文本可读性预测力	82
5.1 中国学者与本族语学者期刊论文句法复杂度差异.....	82
5.2 中国学者期刊论文句法复杂度对文本可读性预测力.....	85

5.2.1 中国学者期刊论文句法复杂度与 FRE 回归分析	85
5.2.2 中国学者期刊论文句法复杂度与 SMOG 回归分析	87
5.3 本族语学者期刊论文句法复杂度对文本可读性预测力	89
5.3.1 本族语学者期刊论文句法复杂度与 FRE 回归分析	89
5.3.2 本族语学者期刊论文句法复杂度与 SMOG 回归分析	91
5.4 中国学者与本族语学者期刊论文句法复杂度与文本可读性; 文本实例	93
5.5 本章小结	98
 第六章 不同学科期刊论文词汇复杂度对文本可读性预测力	101
6.1 不同学科期刊论文词汇复杂度和文本可读性差异	101
6.2 不同学科期刊论文词汇复杂度对文本可读性预测力	112
6.2.1 不同学科期刊论文词汇复杂度与 FRE 回归分析	113
6.2.2 不同学科期刊论文词汇复杂度与 SMOG 回归分析	115
6.3 不同学科期刊论文词汇复杂度与文本可读性: 文本实例 ...	117
6.4 本章小结	121
 第七章 不同学科期刊论文句法复杂度对文本可读性预测力	124
7.1 不同学科期刊论文句法复杂度差异	124
7.2 不同学科期刊论文句法复杂度对文本可读性预测力	130
7.2.1 不同学科期刊论文句法复杂度与 FRE 回归分析	130
7.2.2 不同学科期刊论文句法复杂度与 SMOG 回归分析	132
7.3 不同学科期刊论文句法复杂度与文本可读性: 文本实例 ...	134
7.4 本章小结	136
 第八章 综合讨论	139
8.1 词汇复杂度与文本可读性	139
8.2 句法复杂度与文本可读性	141
8.3 词汇复杂度学科差异与文本可读性	143

8.4	句法复杂度学科差异与文本可读性	144
8.5	词汇复杂度与句法复杂度指标修正	146
8.6	英语通用语视角词汇复杂度、句法复杂度和文本可读性 特征	148
第九章 结语.....		150
9.1	主要发现及意义	150
9.2	研究启示	154
9.3	局限与展望	157
参考文献.....		160

表 目 录

表 3-1	词汇复杂度测量指标	37
表 3-2	句法复杂度测量指标	40
表 3-3	FRE 分数段难度及水平描述	41
表 3-4	多音节单词数量与 SMOG 年级对照表	42
表 3-5	语料信息汇总	44
表 3-6	学术写作元话语分类、功能和实例	49
表 4-1	词汇复杂度、句法复杂度、文本可读性测量数据正态分布 检验	52
表 4-2	中国学者与本族语学者期刊论文词汇复杂度 Mann-Whitney U 检验结果	54
表 4-3	中国学者和英语本族语学者期刊论文文本可读性的比较 ...	58
表 4-4	中国学者期刊论文词汇复杂度与 FRE 多元回归分析	59
表 4-5	中国学者期刊论文词汇复杂度与 SMOG 多元回归分析	63
表 4-6	英语本族语学者期刊论文词汇复杂度与 FRE 多元回归 分析	66
表 4-7	英语本族语学者期刊论文词汇复杂度与 SMOG 多元回归 分析	69
表 4-8	中国学者和本族语学者期刊论文前 500 词(块)元话语	72
表 5-1	中国学者与本族语学者期刊论文句法复杂度 Mann-Whitney U 检验结果	83
表 5-2	中国学者期刊论文句法复杂度与 FRE 多元回归分析	86
表 5-3	中国学者期刊论文句法复杂度与 SMOG 多元回归分析	88

表 5-4 英语本族语学者期刊论文句法复杂度与 FRE 多元回归分析..... 90

表 5-5 英语本族语学者期刊论文句法复杂度与 SMOG 多元回归
分析 92

表 6-1 中国学者期刊论文词汇复杂度 Kruskal-Wallis H 检验结果 102

表 6-2 词汇复杂度四个学科六次事后检验结果 107

表 6-3 中国学者期刊论文四个学科文本可读性 One-Way ANOVA
检验结果 111

表 6-4 理工学科和人文社科学科前 500 词(块)元话语 118

表 7-1 中国学者期刊论文句法复杂度 Kruskal-Wallis H 检验结果 ... 125

表 7-2 句法复杂度四个学科六次事后检验结果 129

图 目 录

图 1-1	研究思路与研究内容章节对应图	9
图 2-1	词汇复杂度、句法复杂度、文本可读性、英语语言能力 关系图	34
图 3-1	词汇复杂度、句法复杂度、文本可读性关系分析过程图	47
图 4-1	中国学者和本族语学者期刊论文文本可读性测量值 FRE 对比图	57
图 4-2	中国学者和本族语学者期刊论文文本可读性测量值 SMOG 对比图	57
图 6-1	中国学者各学科期刊论文文本可读性 FRE 差异	110
图 6-2	中国学者各学科期刊论文文本可读性 SMOG 差异	110
图 6-3	以法律学科为参照, 预测各学科 FRE 与 SMOG 词汇复杂度 特征	122
图 7-1	以法律学科为参照, 预测各学科 FRE 与 SMOG 句法复杂度 特征	137
图 9-1	本研究涌现的研究主题	154
图 9-2	本研究启示	155
图 9-3	语言能力表现形式与深层语言能力	158

第一章 绪 论

1.1 研究背景

期刊论文写作是学术英语写作(English for Academic Writing, 即EAW)研究的焦点之一。长久以来英语一直是学术界公认的国际论文发表语言, 据统计, 美国科学信息研究所(Institute for Scientific Information, 即ISI)数据库中超过90%的知名期刊使用英语作为出版语言(Hyland, 2015), 国际期刊论文也因此成为全球学术界交流学术成果的重要载体(Flowerdew, 1999)。期刊论文写作研究聚焦与学术英语写作相关的英语语言教学与实践, 其主要目的是为了帮助专业人员(包括教研人员、学生、学者等)开展有效的学术研究和交流实践活动。

近年来, 随着宏观层面国家政策影响和微观层面个人发展需要, 越来越多的中国学者不再囿于发表国内期刊论文, 转而发表国际期刊论文。国际期刊论文写作目的是在世界范围内传播新发现与新知识, 实现该目的是以论文能实现作者与读者间的有效沟通为前提。然而, 中国学者们作为英语外语使用者要写出合乎国际期刊规范且表意清晰的英语论文并不容易, 其原因主要体现在以下三个方面。

首先, 期刊论文写作具有复杂性。这种复杂性主要由学术写作语体特殊性决定, 其表现之一是重视词汇复杂度(Lexical Complexity)和句法复杂度(Syntactic Complexity), 即词汇和句法简单被认为是学术英语写作的严重缺陷(Hinkel, 2003)。期刊论文语场(Field)为学术英语领域, 语式(Mode)为正式书面语, 基于此语域(Genre)特征, 期刊论文词汇和

句法也呈现复杂性特征。词汇方面,除了专业技术词汇传递学术信息外,期刊论文还包括非术语的学术词汇或词块用于表达学术观点、阐述研究活动、解释研究结果等,如 AWL(Academic Word List)(Coxhead, 2012)、AFL(Academic Formulas List)(Simpson-Vlach and Ellis, 2010)的创建都表明学术词汇复杂度要求较高。句法方面,期刊论文与其他学术语篇一样惯于使用长句、长 T 单位(T-unit)以及复杂句和复杂结构,如从句、名词化结构等(Brown and Yule, 1983; Halliday and Martin, 1993)。因此,期刊英语论文写作者需要有较丰富的词汇和句法知识储备。

其次,期刊论文写作要求表述详尽(Elaborate)和表意清晰(Explicit)(Hyland, 2002)。期刊论文作为学术界交流的主要媒介,其话语范围包含科技领域知识的传播、延伸与探索。虽然期刊论文的读者是来自同一科研团体的专家、学习者等,但是如果论文语言太晦涩,不仅会造成理解困难(特别是对学术新人而言),同时可能因为理解问题造成错误引用。由于学术英语书面语语篇(包括期刊论文)是脱离语境(Decontextualized)的自主论述,因此写作过程中应保证所有假设、逻辑关系具体而清晰(Johns, 1997)。与学术口语语境不同的是,学术书面语语言更简洁,加之期刊论文篇幅限制,写作者们常通过名词化结构、嵌套从句或短语结构使其表述更具体明确(Hughes, 1996)。因此,期刊英语论文写作者需要具备使用精练语言精准表述的能力。

最后,期刊论文写作的复杂性和详尽清晰表述很难折中。与其他英语非母语学者一样,语言障碍是中国学者发表国际期刊论文难以突破的藩篱,如何恰当使用词汇与句法准确表达自己的观点是中国学者国际期刊英语论文写作拦路虎之一(Flowerdew, 1999; 黄萍、赵冰, 2010)。由于期刊论文词汇复杂度、句法复杂度、可读性关系尚不清楚,学者们在期刊论文写作中常会走向两个极端,一是精简化,即写作中大量使用有限词汇和简单句法(如短句、并列句等),这种做法有助于提高论文可读性,然而有限的词汇可能导致语义表述不完整,简单句法可能使期刊论文这种正式书面语体呈现口语化特征(Hinkel, 2003)。二是复杂

化,即写作中大量使用复杂词汇和句法(如名词化结构、从句等),这有助于呈现期刊论文的书面语复杂性特征,但是大大降低了论文的可读性(Biber and Gray, 2010),容易造成“狐狸博士现象”(Dr. Fox Phenomenon),即用复杂难懂的论文彰显高超的学术水平实则不利于知识传播(Armstrong, 1980)。相关研究表明,期刊论文领域可能已出现“狐狸博士现象”端倪,如Lei和Yan(2016)通过分析期刊论文可读性和引用率关系发现,部分期刊论文可读性越低,其引用率越高。

鉴于期刊论文词汇复杂度、句法复杂度、可读性关系尚不明确的现状,以及平衡三者关系对期刊论文写作的必要性,本研究试图从词汇复杂度、句法复杂度、文本可读性等方面描述中国学者国际期刊论文语言复杂度特征,并进一步构建词汇复杂度和句法复杂度对文本可读性预测力理论模型,便是本研究的动机之一。

词汇复杂度和句法复杂度源于Larsen-Freeman(1997)的复杂理论,该理论为解释语言和语言习得现象提供了方法论原则。语言复杂度(Complexity)、准确度(Accuracy)和流利度(Fluency),即CAF,在语言和语言习得研究中常用于衡量静态的语言产出水平或描述动态的语言产出发展轨迹和规律(如,Skehan, 2009;江韦姗、王同顺, 2015;郑咏滢, 2011等)。此外,语言产出的复杂度、准确度和流利度特征也用于衡量语言产出质量,反映语言使用者使用语言知识储备陈述个人观点的能力(如,Beers and Nagy, 2009; Crowhurst, 1983; Gregg et al., 2000; Ortega, 2003; Yang et al., 2015等)。语言产出复杂度(包括词汇复杂度和句法复杂度)是书面语研究中最常关注的问题(Wolfe-Quintero et al., 1998; 鲍贵, 2004, 2009; 江韦姗、王同顺, 2015)。横向比较研究表明,不同水平和不同母语背景的英语语言使用者书面语语言复杂度存在差异,如鲍贵(2009)选取T单位平均长度、子句平均长度、子句从属句比率等多项指标观测不同年级英语学习者写作句法复杂度,发现T单位平均长度和子句平均长度两项指标在不同年级间有显著差异。又如,Lu和Ai(2013; 2015)研究发现,英语本族语和非本族语者之间,以及不同文化背景非本族语者之间,英语写作中句法复杂度存在差异。

纵向历时研究表明,语言产出复杂度在不同方面均呈非线性动态发展,即二语学习者的语言产出复杂度在语言发展不同阶段呈现不同特征,如郑咏滢(2015b)通过分析一学年间16名英语专业学生作文词汇复杂度指标(包括词汇复杂程度、词汇多样性和词汇密度)发现,词汇复杂程度发展基本处于停滞状态,但词汇多样性和词汇密度在波动中显著上升。

语言复杂度可以从词汇和句法层面直观衡量语言产出质量,因而广泛运用于二语学习者语言水平发展研究,但研究表明语言复杂度与语言产出阅读理解也息息相关(Metoyer-Duran, 1993; Rottensteiner, 2010)。书面语是语言使用者运用词汇和句法知识储备陈述观点的重要传输介质,其文本可读性(Readability)直接决定语言使用者的观点是否被准确传递与理解,因此,期刊论文写作过程中如果因词汇复杂度或句法复杂降低论文可读性可能导致论文发表困难(Flowerdew, 1999; 黄萍、赵冰, 2010),毕竟没有人愿意去读他们读不懂的文章(Otto et al., 2011)。

Flowerdew 等人曾就中国学者(特别是香港地区学者)发表国际期刊论文问题开展一系列研究,他们发现除了专业知识外,中国学者面临的问题之一即如何使用恰当的英语词汇和句法撰写国际期刊论文(Flowerdew, 1999, 2008; Flowerdew and Li, 2009; Flowerdew and Peacock, 2001)。另一方面,基于期刊编辑和审稿人视角研究也发现恰当词汇和句法结构的使用在一定程度上能增强文本可读性,更有利于编辑、审稿人以及目标读者理解论文内容,为论文最终录用与发表增加砝码(Coleman, 2014; Coniam, 2012)。这些研究均表明,词汇和句法结构对期刊论文写作的影响已引起语言研究领域学者的关注,然而,由于词汇和句法测量指标较多,且学者们尚未对权威的词汇和句法结构指标达成一致意见(Anderson, 2012; Ortega, 2003),因此很少有人研究词汇和句法结构对文本可读性影响的具体指标和机制。

文本可读性研究关注文本表述风格,即词汇和句子结构对读者阅读理解文本的影响(Dale and Chall, 1948)。由于文本可读性在衡量文本

阅读难度和甄选设计文本材料(如教材、宣传手册等)过程中起到关键作用,研究者们尝试用多种方法测量文本可读性(DuBay, 2004; Flesch, 1948; Klare, 1963; McLaughlin, 1969)。目前,科技论文文本可读性测量指标主要为 FRE(Flesch Reading Ease)和 SMOG(Simple Measure of Gobbledygook),前者采用所有单词数占有所有句子数比率和所有音节数占有所有单词数比率测量文本阅读理解难度,后者从多音节单词数量上进行考量。可以看出,这两种方法均停留在单词和音节数量层面,缺乏考量句法结构层面对文本可读性的影响。作为文本的筋骨,有必要从词汇和句法两方面探究影响文本可读性因素,以完善和提高文本可读性测量方法。

因此,本研究通过词汇复杂度和句法复杂度对文本可读性预测力模型分析,探究词汇和句法维度影响文本可读性的具体指标。此外,由于语言和文化背景可能影响本族语和非本族语者的期刊论文写作,且各学科书面语写作中短语和句子复杂度存在差异(Staples et al., 2016),本研究还开展中国学者与英语本族语学者,以及不同学科中国学者两项对比研究,以便于从词汇复杂度和句法复杂度层面总结中国学者期刊论文写作语言复杂度特征和学科特色。

1.2 研究问题

本书从词汇复杂度、句法复杂度以及文本可读性三方面对期刊论文进行分析,以发表的国际期刊论文语料作为研究对象,通过计算和整合文本词汇复杂度、句法复杂度、文本可读性三方数据,以建立不同条件下词汇复杂度和句法复杂度对文本可读性预测力模型。

本书主要关注层面是词汇复杂度和句法复杂度对文本可读性的影响与预测,同时也关注中国学者相较于本族语者期刊论文所特有的词汇和句法特征,以及不同学科期刊论文词汇和句法特征。我们首先对比中国学者和本族语学者期刊论文词汇复杂度、句法复杂度以及文本可读性,基于对比结果,建立预测力模型,以找出中国学者期刊论文词汇和句法