

中外大学生英语作文中的词汇和句法 复杂度对比研究

潘玉杰, 鲍 贵

(南京工业大学 外国语学院, 江苏 南京 211816)

摘要: 通过对比分析英语专业4个年级学习者作文和英语母语者作文, 从词汇复杂度和句法复杂度角度, 探究中国英语专业学生的书面语变化特点以及词汇和句法复杂度之间的关系。词汇复杂度的测量指标为词汇多样性和词汇复杂性, 句法复杂度的测量指标为T单位长度、T单位复杂性比率和子句中的名词短语数。单因素多元方差分析发现: 英语专业学习者词汇和句法复杂度各指标随年级的提高而提高; 除了词汇多样性和T单位复杂性比率无显著性差异外, 其余指标均显著低于母语者; 词汇复杂度与句法复杂度除个别指标外, 其余指标相互独立, 无显著相关性。该结果对深入研究二语写作教学和指导教学实践有重要启示。

关键词: 英语专业学生; 词汇复杂度; 句法复杂度; 二语写作

中图分类号: H319 **文献标识码:** A **文章编号:** 1009-895X(2026)01-0028-09

DOI: 10.13256/j.cnki.jusst.sse.240625309

A Contrastive Study of Lexical and Syntactic Complexity in English Compositions by Chinese and Foreign College Students

PAN Yujie, BAO Gui

(School of Foreign Languages and Literature, Nanjing Tech University, Nanjing 211816, China)

Abstract: This study presents a contrastive analysis on compositions by Chinese English majors across four grades and those of native speakers. It investigates the characteristics of written language in Chinese English majors in terms of lexical complexity and syntactic complexity, exploring the mutual relationship between the two complexities. Lexical complexity is measured by lexical diversity and lexical sophistication, while syntactic complexity is assessed through T-unit length, T-unit complexity ratio, and complex nominals per clause. One-way multivariate analysis of variance (MANOVA) reveals that the measures of lexical and syntactic complexity improve gradually from grade 1 to grade 4. Compared with native speakers, Chinese English majors demonstrate significantly lower measures, except for lexical diversity and the T-unit complexity ratio, which are similar to those of native speakers. The measures of lexical and syntactic complexity are largely independent of each other, with few significant correlations observed. This study provides important insights for further research on second language writing

收稿日期: 2024-06-25

基金项目: 江苏省社会科学基金项目(23YYB008)

第一作者: 潘玉杰, 女, 硕士研究生。研究方向: 二语习得。E-mail: 1433266563@qq.com

通信作者: 鲍 贵, 男, 教授。研究方向: 二语习得。E-mail: boggy2008@163.com

instruction in practice.

Keywords: English majors; lexical complexity; syntactic complexity; second language writing

语言复杂度一直是二语写作研究的中心议题,通常包括词汇复杂度和句法复杂度两大方面,是衡量学习者语言发展的重要维度,能有效预测写作质量^[1-3]。词汇复杂度,又称词汇丰富性,指词汇的多样性和复杂性;句法复杂度指语言结构的多样性和复杂化程度^[4]。现有研究大多集中在词汇或句法单一维度,考察相对片面。有学者指出,短语复杂度的测量也应纳入句法复杂度的考察范围^[5-7]。此外, Norris 等也指出,研究者应重视语言复杂度子成分的发展途径,并关注它们之间的关联性^[8]。鉴于此,本研究结合短语复杂度,对比分析中国英语专业学生和英语本族语者的书面语材料,以年级水平为自变量,考察词汇和句法复杂度的多项指标以及词汇复杂度与句法复杂度之间的互动关系,探究中国英语专业学生在英语写作中的词汇和句法复杂度特点,为英语写作教学提供参考。

一、文献综述

关于词汇和句法复杂度的前期研究大致可以分为以下三类。第一类是研究词汇或句法复杂度与二语写作质量之间的关系^[9-12]。此类研究发现词汇复杂度和句法复杂度与二语写作质量之间存在相关性,但在哪些指标能够更好地衡量二语写作质量方面存在差异。例如,在句法复杂度研究中, Casal 等的研究发现, T 单位长度、T 单位复杂性比率和子句中名词短语数量可以有效预测二语写作质量,但是从句复杂度与二语写作质量之间不存在显著相关性^[6]。高霞发现宏观句法复杂度指标比微观句法复杂度更能有效预测写作质量^[12]。在词汇复杂度与二语写作质量的关系研究方面,王海华等发现词汇多样性、词汇复杂性和词汇密度与写作质量显著相关,词汇错误和写作质量之间呈显著负相关^[11]。但 Lemmouh 的研究发现词汇丰富性与二语写作质量之间没有相关性^[9]。

第二类是二语学习者词汇复杂度和句法复杂度的发展特征。该类研究通常对某一群体进行跟踪考察^[13-16]。Ishikawa 发现,随着二语学习者水

平的提高,学习者的词汇多样性先下降后上升,词汇密度保持不变,词汇复杂性逐步上升^[15]。Bulté 等对 45 名英语学习者的词汇和句法复杂度进行一学期的跟踪调查,结果发现大多数句法复杂度指标经过一段时间的学习后显著增长,词汇多样性没有改变,词汇复杂性同样显著增长^[13]。国内学者朱慧敏等对 30 名英语专业学生的同题议论文进行为期 4 年的历时考察,发现在英语水平提高的过程中,词汇多样性变化呈现波动性,词汇密度和词汇复杂性呈上升态势^[16]。

最后一类研究聚焦于词汇复杂度与句法复杂度对比。该类研究通常对比不同年级或不同水平的词汇复杂度与句法复杂度的差异^[17-20]。例如, McNamara 等利用 Coh-Metrix 分析了 120 篇英语母语者的作文,结果发现高水平作文的词汇多样性和词汇复杂性均比低水平作文高^[17]。Martínez 探讨了学习者在三、四年级的句法使用特征变化,发现他们几乎在所有句法复杂度指标上都呈显著增长^[18]。张会平对英语初学者的词汇丰富性分析表明,学习者的词汇复杂度、词汇多样性和词汇密度随年级的提高而提高,但各维度发展不均衡^[3]。也有研究者对比分析不同语言背景的词汇或句法复杂度。例如, Linnarud 对比分析瑞典学生与英语本族语者作文中词汇丰富性的差异,结果发现在词汇错误、词汇多样性、词汇新颖性、词汇搭配和个性化词汇使用方面,二者有显著差异^[21],本族语者的学生作文显著优于瑞典英语学习者。雷蕾对比研究了英语学习者和母语者硕博士论文中的句法复杂度,结果发现,中国英语学习者的语言产出长度和从属结构与本族语者有显著差异,名词短语复杂度和并列结构数量与本族语者无显著差异^[22]。整体上,学习者的语言复杂度随着年级或英语水平的提高而提高,但与母语者相比尚有一定差距。

综上,虽然以往研究在学习者作文词汇和句法复杂度方面取得了显著成绩,得出一些重要的结论,但尚存在不足。首先,前人的研究多为单独考察词汇和句法的变化趋势,没有将词汇和句法相结合考察学习者书面语言变化的特点。其次,

目前的研究多忽视了词汇复杂度和句法复杂度之间存在何种互动关系的问题。最后,在统计方法上,目前的研究多使用单因素方差分析,易扩大第一类错误率,鲜有研究使用多元统计分析。鉴于此,本研究参照母语者的书面作文语料,采用多元统计方法,探究大学一至四年级英语专业学习者词汇复杂度和句法复杂度的变化特点以及二者之间的互动关系,从而全面考察书面语作文语言复杂度的变化特点。

二、研究设计

(一) 研究问题

本研究主要回答以下问题:

(1) 学习者词汇复杂度、句法复杂度随年级水平提升呈现怎样的变化趋势?与母语者相比特点如何?

(2) 词汇复杂度和句法复杂度之间存在何种关系模式?

(二) 研究对象

本研究作文语料来源于中国大学生英语笔语语料库(WECCL)和鲁汶大学的LOCNESS语料库。中国大学生英语笔语语料库由文秋芳教授主持创建^[23],是中国学生英语口笔语语料库中的子库,容量为100万词。本研究的语料来自于WECCL语料库中英语专业一至四年级学生的议论文,每个年级选取40篇,共计160篇。综合学习者对话题的熟悉度,各年级题目统一为“Nowadays, more and more college students rent apartments and live outside campus. Is it appropriate? State your opinion about this.”要求限时写作,长度为300词左右。

LOCNESS语料库由鲁汶大学建立,容量大约为30万词,包含英美母语者高中生和大学生的书面语材料,是本研究的参照语料库。为使体裁和作文时间相近,本研究选取由美国密歇根大学的大学生撰写的40篇议论文,学生年龄与中国学生相当,大致分布在19~23岁。作文题目为“Great inventions and discoveries of 20th century and their impact on people's lives.”字数要求为不超过500个词,限时完成,作文基本数据见表1。

表1 不同年级英语学习者和母语者作文情况描述性统计
Tab.1 Descriptive statistics for the compositions of EFL learners at different grades and native speakers

研究对象	作文数量	平均长度	最大值	最小值	标准差
一年级	40	358	547	176	74.73
二年级	40	352	580	190	97.36
三年级	40	373	597	220	79.11
四年级	40	314	434	145	62.20
母语者	40	378	568	214	84.01

表1显示,从作文长度看,母语者作文的平均长度高于英语专业学习者。各个年级作文的平均长度在300~400词之间,4个年级中三年级学习者的作文平均长度最长,接近母语者作文的平均长度,四年级作文平均长度最低。

(三) 变量操作定义与测量

本研究中词汇复杂度测量采用词汇多样性和词汇复杂性两个测量指标。词汇复杂性(lexical sophistication),即文中低频词占总词数的比率,其计算方式为低频词类符数与文本总字数的比值,再乘以100。本研究参照Lu^[24]的做法,将低频词定义为BNC(英国国家语料库)词表1和词表2以外的词。词汇多样性(lexical diversity),指文本中类符与形符的比值(type-token ratio, TTR)。但是TTR易受到文本长度影响,移动平均类符形符比(MATTR)受文本长度的影响很小^[25],计算方法为按照固定的窗口(词符数),从文本开头计算TTR,然后在文本中依次移动窗口计算TTR,直至在包括文本最后一个词符的窗口中计算TTR,最后计算这些TTR的平均值。因此,本研究选用MATTR来测量词汇多样性。词汇多样性的测量工具为Michalke开发的R数据包koRpus^[26]。

句法复杂度测量根据Lu^[1]和鲍贵^[20]的研究,采用3个测量指标,即:T单位长度(MLT)、T单位复杂性比率(C/T)和子句中的名词短语数(CN/C)。T单位指一个主句或由一个主句加一个或多个从属句组成。T单位长度指每个文本的总词数除以该文本包含的T单位数(MLT)。T单位复杂性比率指每个T单位中包含的子句数(C/T)。子句中的名词短语数(CN/C)指子句中复杂名词短语的平均值。3个指标采用Tan开发的句法分析器NeoSCA进行测量,NeoSCA是

L2SCA 的重写版本可在 Windows 系统下使用^[27]。

(四) 统计分析

本研究为被试间设计,对不同年级学习者和母语者作文中的词汇和句法复杂度进行正态性检验发现,有19个单元格数据违反正态分布假设。譬如,二年级的词汇复杂性数据违反正态分布($W=0.87, p=0.00<0.05$)。利用中位数的 Levene 方差齐性检验发现,不同年级水平上的 T 单位长度(C/T)数据不满足方差齐性假设($F_{(3,156)}=3.46, p=0.02<0.05$)。

鉴于多组数据违反正态性假设和方差齐性假设,本研究采用稳健型多元统计方法,这种统计方法不要求数据满足正态检验和方差齐性假设,统计分析软件为 R 4.2.1,稳健多元方差分析调用 R 数据包 Rallfun-v40 中的 MULAOPp() 函数,对单因变量的测量利用该数据包中的函数 tlway(), 采用 20% 截尾平均数的 Welch 型方法进行单因变量方差分析,设定显著性水平为 $\alpha=0.05$ ^[28]。配对比较效应量估计采用 Cohen's d_R , 效应量大小采用 Cohen 建议的标准: $d_R=0.2$, 效应量小; $d_R=0.5$, 效应量中等; $d_R=0.8$, 效应量大^[29]。

三、研究结果

(一) 学习者词汇和句法复杂度的变化

1. 词汇复杂度

本节报告词汇复杂度各测量指标随年级提高的变化特点。表2显示词汇复杂度测量指标的描述性统计结果,包括20%截尾平均数($\bar{X}_t$)、20%缩

尾标准差(S_w)和平均数95%置信区间(95%CI)。

表2显示:各年级在词汇多样性测量指标上的平均值接近一致,整体变化趋势不明显;词汇复杂性则逐年上升,呈由低到高的增长趋势。

单因素多元稳健方差分析发现,词汇复杂度在4个年级中存在显著性差异($Q=21.85, p=0.010<0.050$)。进一步的单因变量方差分析表明,词汇复杂性在4个年级中存在统计显著性差异($F(3, 50.85)=6.30, p<0.001$),词汇多样性不存在统计显著性差异($F(3, 51.07)=0.75, p=0.520>0.050$)。在词汇复杂性测量指标上的配对比较结果显示,二年级的词汇复杂性显著好于一年级,且效应量中等($p=0.040<0.050, d_R=0.50$)。三年级和四年级的词汇复杂性也显著好于一年级的词汇复杂性(前者: $d_R=0.78, p<0.050$; 后者: $d_R=0.98, p<0.050$),效应量分别处于中等和较高水平。其他年级配对比较均没有显著差异,效应量处于较低水平。

为考察与母语者的差距如何,对词汇复杂度指标在不同年级学习者和母语者之间的差异进行比较,结果见表3。

表3显示,一至四年级学生的词汇多样性与母语者基本持平,词汇复杂性则均低于母语者。配对比较结果显示,词汇多样性与母语者之间不存在统计显著性差异($p>0.050$),效应量均处于较低水平,词汇复杂性均显著低于母语者($p<0.001$),效应量大。

2. 句法复杂度

本节报告句法复杂度各测量指标随年级提高的变化特点。句法复杂度指标的描述性统计结果如表4所示,包含20%截尾平均数($\bar{X}_t$)、20%缩尾

表2 不同年级英语学习者和母语者作文中的词汇复杂度测量指标描述性统计

Tab. 2 Descriptive statistics of lexical complexity measures for the compositions of EFL learners at different grades and native speakers

研究对象	词汇多样性			词汇复杂性		
	$\bar{X}_t$	S_w	95%CI	$\bar{X}_t$	S_w	95%CI
一年级	0.70	0.02	[0.69, 0.71]	3.30	0.83	[2.85, 3.75]
二年级	0.71	0.02	[0.69, 0.72]	4.25	1.08	[3.46, 4.64]
三年级	0.70	0.02	[0.68, 0.71]	4.39	0.96	[3.86, 4.91]
四年级	0.70	0.02	[0.68, 0.71]	4.72	1.03	[4.16, 5.29]
母语者	0.71	0.03	[0.69, 0.72]	6.21	1.47	[5.41, 7.01]

表3 词汇复杂度指标在不同年级英语学习者和母语者作文中的差异比较

Tab. 3 Comparison of lexical complexity measures between the compositions of EFL learners at different grades and native speakers

对象对比	词汇多样性					词汇复杂性				
	平均数差异	SE	t	p	d _R	平均数差异	SE	t	p	d _R
一年级-母语者	-0.01	0.01	0.62	0.54	-0.15	-3.28	0.45	6.50	0.00	-1.57
二年级-母语者	0.00	0.01	0.09	0.93	0.02	-2.25	0.49	4.46	0.00	-1.08
三年级-母语者	-0.01	0.01	0.94	0.35	-0.23	-1.93	0.47	3.92	0.00	-0.95
四年级-母语者	-0.01	0.01	1.26	0.22	-0.30	-1.70	0.48	3.11	0.00	-0.75

表4 不同年级英语学习者和母语者作文中的句法复杂度测量指标描述性统计

Tab. 4 Descriptive statistics of syntactic complexity measures for the compositions of EFL learners at different grades and native speakers

研究对象	T单位长度			T单位复杂性比率			子句中名词短语数		
	$\bar{X}_t$	S_w	95%CI	$\bar{X}_t$	S_w	95%CI	$\bar{X}_t$	S_w	95%CI
一年级	13.05	1.50	[12.23, 13.86]	1.52	0.09	[1.47, 1.57]	0.73	0.13	[0.66, 0.81]
二年级	13.51	1.61	[12.64, 14.39]	1.49	0.14	[1.42, 1.56]	0.80	0.13	[0.73, 0.87]
三年级	14.76	2.21	[13.55, 15.96]	1.61	0.20	[1.49, 1.72]	0.89	0.18	[0.79, 0.98]
四年级	15.98	2.01	[14.89, 17.07]	1.62	0.19	[1.52, 1.73]	0.92	0.17	[0.83, 1.01]
母语者	16.05	1.88	[15.02, 17.08]	1.57	0.14	[1.50, 1.65]	1.18	0.17	[1.08, 1.27]

标准差 (S_w) 和平均数 95% 置信区间 (95%CI)。

从表4的数据可知, 随年级水平升高, T 单位长度 (MLT) 和每个子句中名词短语数量 (CN/C) 呈直线上升趋势, 而 T 单位复杂性比率 (C/T) 在一至二年级阶段略有下降, 二至四年级阶段呈小幅上升趋势。

单因素多元方差分析发现, 句法复杂度测量指标在 4 个年级中存在显著性差异 ($Q=29.19, p=0.01<0.05$)。单因变量方差分析结果显示, T 单位复杂性比率 (C/T) 不存在统计显著性差异 ($F_{(3, 48.81)}=2.05, p=0.12>0.05$), T 单位长度 (MLT) 和子句中名词短语数量 (CN/C) 在 4 个年级中存在统计显著性差异 (MLT: $F_{(3, 50.85)}=7.30, p<0.001$; CN/C: $F_{(3, 50.65)}=4.39, p=0.01<0.05$)。在 T 单位长度 (MLT) 上的配对检验发现, 三、四年级组均显著好于一年级组 ($p<0.05$), 稳健效应量为 0.58 和 1.06, 分别处于中等和较高水平; 四年级的 T 单位长度 (MLT) 也显著高于二年级, 且效应量大 ($p<0.001, d_R=0.87$)。其他所有配对比较没有显著性差异, 效应量处于较低或者中等水平。

每个子句中的名词短语数量 (CN/C) 上的配对检验显示, 三、四年级的子句中名词短语数量 (CN/C) 显著好于一年级组 ($p<0.05$), 稳健效应量分别为 0.64 和 0.79, 均处于中等水平; 四年级组的子句中名词短语数量也显著高于二年级组, 效应量为中等水平 ($p=0.03<0.05, d_R=0.53$)。其他年级的配对比较不存在显著性差异, 效应量处于中等或较低水平。

由表4数据可知, 一至四年级学生的 T 单位长度 (MLT) 和子句中名词短语数 (CN/C) 均低于母语者; T 单位复杂性比率 (C/T) 在一、二年级阶段低于母语者, 三、四年级阶段高于母语者。为检验与母语者之间是否存在显著性差异, 表5对句法复杂度指标在不同年级学习者和母语者作文中的差异进行报告。

表5显示, T 单位长度 (MLT) 在一、二年级阶段显著低于母语者 ($p=0.00<0.001$), 效应量较大; 三、四年级阶段则不存在统计显著性差异 ($p>0.05$), 效应量处于较小或中等水平。但是, 一至四年级学生的子句中名词短语数量 (CN/C) 则均显著低于母语者 ($p<0.001$), 效应量处于较

表5 句法复杂度指标在不同年级英语学习者和母语者作文中的差异比较

Tab. 5 Comparison of syntactic complexity measures between the compositions of EFL learners at different grades and native speakers

对象对比	T单位长度				T单位复杂性比率				子句中名词短语数			
	平均数差异	SE	p	d _r	平均数差异	SE	p	d _r	平均数差异	SE	p	d _r
一年级—母语者	-0.01	0.64	0.00	-1.13	-3.28	0.45	0.24	-0.29	-0.50	0.06	0.00	-1.83
二年级—母语者	0.00	0.66	0.00	-0.93	-2.25	0.49	0.12	-0.38	-0.42	0.06	0.00	-1.60
三年级—母语者	-0.01	0.77	0.35	-0.41	-1.93	0.47	0.61	0.12	-0.33	0.07	0.00	-1.06
四年级—母语者	-0.01	0.73	0.22	-0.02	-1.70	0.48	0.46	0.18	-0.27	0.07	0.00	-0.94

高水平。各年级的T单位复杂性比率(C/T)与母语者之间均不存在显著性差异($p > 0.05$),效应量则处于中等或较低水平。

(二) 词汇和句法复杂度之间的关系模式

问题2 探讨词汇和句法复杂度之间的关系模

式,对每个年级词汇和句法复杂度指标之间的稳健相关系数进行计算,结果见表6。左面表格对角线下为一年级的稳健相关系数矩阵,对角线上为二年级的稳健相关系数矩阵;右面表格对角线下为三年级的稳健相关系数矩阵,对角线上为四年级的稳健相关系数矩阵。

表6 不同年级水平各测量指标之间的稳健相关系数矩阵

Tab. 6 Robust correlation matrix among the measures of lexical complexity and syntactic complexity at different grades

测量指标	1	2	3	4	5	测量指标	1	2	3	4	5
1 LD	1	-0.02	0.07	0.12	0.04	1 LD	1	-0.09	0.28	0.20	0.34*
2 LS	0.12	1	0.33	-0.03	0.50*	2 LS	0.11	1	0.02	-0.15	0.30
3 MLT	0.24	0.16	1	0.58**	0.61**	3 MLT	-0.01	0.26	1	0.77**	0.45**
4 C/T	0.24	0.08	0.64**	1	0.18	4 C/T	-0.06	0.23	0.81**	1	-0.14
5 CN/C	-0.01	0.23	0.60**	0.25	1	5 CN/C	-0.03	0.07	0.59**	0.31*	1

注: *和**分别表示在0.05和0.01水平上显著相关。

表6显示,在各年级水平上,词汇复杂度两项指标(词汇复杂性和词汇多样性)均呈低相关,未达到统计显著性差异($p > 0.05$)。在句法复杂度三个测量指标方面,各个年级的T单位长度(MLT)与T单位复杂性比率(C/T)和每个子句名词短语数量(CN/C)呈效应量较大或中等水平的显著正相关($p < 0.01$)。T单位复杂性比率(C/T)和子句中名词短语数量(CN/C)在三年级水平上呈中等水平的显著正相关($r = 0.31, p < 0.05$),其他年级之间均没有显著相关性($p > 0.05$)。

在两类测量之间,词汇复杂度各个测量指标与T单位长度(MLT)和T单位复杂性比率(C/T)在4个年级水平上均呈低相关,且没有达到统计显著性($p > 0.05$)。但是,词汇复杂性在二年级水平上与子句中名词短语数量(CN/C)呈高度的显著正相关($r = 0.50, p < 0.01$),词汇多样性在

四年级水平上与子句中名词短语数量(CN/C)呈中等水平的显著正相关($r = 0.34, p < 0.05$)。其他年级词汇复杂度的测量指标与子句中名词短语数量(CN/C)的配对之间均没有显著相关($p > 0.05$)。

以上结果表明,随年级升高,除个别指标外,词汇和句法复杂度各个测量指标之间并未产生显著性关联,说明词汇复杂度和句法复杂度的大部分测量指标之间相互独立,无明显交集。

四、讨论

(一) 学习者词汇和句法复杂度的变化

1. 词汇复杂度的变化

本研究发现词汇多样性随年级升高虽略有波

动,但整体趋于一致,4个年级之间没有显著性差异,说明学习者在语言输出过程中进步缓慢,不擅长使用多样化词汇,这一研究结果与 Laufer^[30]的结论一致。与母语者对比发现,词汇多样性与母语者不存在显著性差异,研究结果与 Linnarud 的结论不同:该研究表明母语者与非母语者学生在词汇多样性方面存在显著差异,造成这种差异的原因可能是语言背景不同^[21]。在 Linnarud 的研究中,研究对象的语言背景为瑞典语,瑞典语和英语同属于日耳曼语系,相似的语言背景可能会促使学习者在写作过程中使用两种语言中的相似词汇,限制词汇多样性的发展;而本研究的研究对象为汉语,属于汉藏语系,语言背景的差异性可能会促使学习者更加注重语言输出过程中词汇的准确使用和多样化表达,从而提高词汇多样性,接近母语者水平。此外,语言水平也可能是导致结论不一的原因。Linnarud 的研究对象为 17 岁的高中生,处于语言学习的早期阶段,语言学习尚不够全面,故其词汇多样性要逊色于母语者;而本研究的研究对象为英语专业的大学生,接受了系统的语言学习,英语水平要优于高中生,因此其词汇多样性会接近母语者。

相对于词汇多样性,学习者书面语产出中的词汇复杂性指标随年级上升而明显提高,且在 4 个年级之间存在显著性差异,与前人研究大体相符。这进一步印证了学习者的词汇复杂性随着学习的深入而增加^[11]。对比结果表明,中国英语专业学生的词汇复杂性显著低于母语者。该发现与 Lei 的结论一致:中国博士生的词汇复杂性低于母语者^[31]。因此,本研究进一步表明中国英语学习者用词稍显简单,语言输出过程中倾向于使用高频词,忽略低频词的使用。其原因可能是由于学习者在词汇学习过程中,仅仅是机械记忆,未能理解词汇的用法。因此,在产出词汇时,学习者往往使用高频词,对于低频词可能因掌握不全而回避使用。

2. 句法复杂度的变化

本研究发现句法复杂度的 3 个测量指标均随年级升高呈直线上升趋势,潜在原因可能是,随着语言水平的提高,学生能够使用各类从句、名词短语和嵌套结构。因此,句法复杂度各测量指标均有不同程度的增加,与 Martínez 的研究结果

一致^[18]。在 4 个年级组的比较中,T 单位长度 (MLT) 和子句中名词短语数量 (CN/C) 两个指标在各年级组之间存在显著性差异,T 单位复杂性比率 (C/T) 在各年级组之间没有显著性差异,这一发现印证了 Lu 的研究,即 C/T 指标无法区分年级水平,MLT 和 CN/C 指标可以区分年级水平^[1]。类似的结果也见于鲍贵对英语学习者作文句法复杂性的变化研究:各年级组在 C/T 指标上没有显著性差异^[20]。这说明学习者的 MLT 和 CN/C 存在明显进步,但是 C/T 无明显发展迹象。

与母语者对比发现,4 个年级组的子句中名词短语数量 (CN/C) 均显著低于母语者,一二年级组的 T 单位长度 (MLT) 指标也显著低于母语者,三四年级组与母语者之间无显著性差异,与 Mancilla 等的研究结论^[32]一致。各年级组的 T 单位复杂性比率 (C/T) 则与母语者之间均不存在显著性差异,这一结果与徐晓燕等的研究结论^[33]不同。徐晓燕等发现 C/T 值显著低于母语者,因为中国大学生偏爱使用简单句并且对句子类型控制不当,从而削弱了 C/T 值^[33]。但是本研究中的 C/T 值与母语者之间无显著性差异。造成这一差异的原因可能是语料来源不同。徐晓燕等的研究中母语者语料来源于国外外语教学类期刊文章,文章水平较高,包含更多的子句,所以 C/T 值显著高于中国大学生^[33]。总之,学习者在书面语产出过程中的句法复杂度仍与母语者存在显著性差距。

(二) 词汇和句法复杂度的关系模式

本研究发现词汇复杂度和句法复杂度仅在个别测量指标之间存在显著关联,大多数指标相互独立,无显著相关性。由此可以推断,学习者的书面语处于相对稳定的状态。该研究结果与郑咏滢等的研究既有相似之处,但也有不同。郑咏滢等认为词汇复杂度和句法复杂度之间既相互独立又相互关联^[34]。然而,本研究认为词汇复杂度和句法复杂度的大部分指标之间并无明显互动,不存在显著相关性,似乎可视为语言复杂度下两个独立的子系统。究其原因,主要是因为郑咏滢等的研究为历时研究,对 15 名英语专业大一学生一学年间词汇复杂度和句法复杂度之间的复杂交互进行考察,结果表明部分学习者词汇和句法复杂度处于相对稳定的状态,两者间无明显关联,但也有部分学习者词汇和句法复杂度相互竞争^[34]。

但是本研究为共时研究,未对同一批学习者进行多次跟踪研究,因此造成研究结果略有不同。

造成研究结果不同的另一个原因可能是本研究为群体研究,因此未能整体观察到词汇复杂度和句法复杂度之间的显著关联。虽然郑咏滢等的研究也为群体研究,但其样本量较少且深入考察个体发展模式,计算了15名学生各自的词汇复杂度指标和句法复杂度指标之间的相关系数,由此得到两类研究结果:第一类为词汇复杂度和句法复杂度高度相关;第二类为词汇复杂度和句法复杂度之间相互独立^[34]。本研究仅对群体进行考察,未进行个案考察,因此结果不同。

五、结语

本研究从词汇复杂度和句法复杂度两个维度,参照母语者语料库,考察了中国英语专业学生大学4年的书面语变化特点以及词汇复杂度与句法复杂度之间的互动关系。研究发现:(1)词汇复杂度和句法复杂度各指标随着学习者年级的升高而上升,与母语者对比发现,词汇多样性和T单位复杂性比率无显著差异,词汇复杂性、子句中名词短语数量和T单位长度差异显著;(2)除部分指标外,词汇复杂度和句法复杂度的大多数测量指标之间不存在显著相关性,无明显交集。

基于以上发现,本研究对二语写作教学提出几点建议。首先,在写作任务开始之前,教师可通过“思维导图”或“头脑风暴”等方式,鼓励学生分享相同意义的不同词汇,引导他们在写作中使用多样化词汇。其次,教师可在教学过程中逐步增加低频词的讲授,同时通过词汇替换、造句、写作等练习,强化学生对低频词的运用,提高词汇复杂性。此外,教师还应引导学生在写作中避免使用简单句,增加从句和各类名词短语修饰语(如定语、形容词短语、介词短语等)的使用,从而增加句子长度和名词短语数量,以提升句式的灵活性和复杂性。最后,学习者也可多接触英语本族语者语料,感受词汇在不同语境下的细微差别,了解复杂句式结构是如何构建的,进而在写作中进行模仿和应用。

参考文献:

- [1] LU X F. A corpus-based evaluation of syntactic complexity measures as indices of college-level ESL writers' language development[J]. *TESOL Quarterly*, 2011, 45(1): 36 – 62.
- [2] BI P, JIANG J Y. Syntactic complexity in assessing young adolescent EFL learners' writings: syntactic elaboration and diversity[J]. *System*, 2020, 91: 102248.
- [3] 张会平. 中国英语初学者写作词汇丰富性的发展特征研究[J]. *现代外语*, 2020, 43(4): 529 – 540.
- [4] CROSSLEY S A, MCNAMARA D S. Does writing development equal writing quality? A computational investigation of syntactic complexity in L2 learners[J]. *Journal of Second Language Writing*, 2014, 26: 66 – 79.
- [5] BIBER D, GRAY B, POONPON K. Should we use characteristics of conversation to measure grammatical complexity in L2 writing development?[J]. *TESOL Quarterly*, 2011, 45(1): 5 – 35.
- [6] CASAL J E, LEE J J. Syntactic complexity and writing quality in assessed first-year L2 writing[J]. *Journal of Second Language Writing*, 2019, 44: 51 – 62.
- [7] LAN G, ZHANG Q S, LUCAS K, et al. A corpus-based investigation on noun phrase complexity in L1 and L2 English writing[J]. *English for Specific Purposes*, 2022, 67: 4 – 17.
- [8] NORRIS J M, ORTEGA L. Towards an organic approach to investigating CAF in instructed SLA: the case of complexity[J]. *Applied Linguistics*, 2009, 30(4): 555 – 578.
- [9] LEMMOUH Z. The relationship among vocabulary knowledge, academic achievement and the lexical richness in writing in Swedish university students of English[D]. Stockholm: Stockholm University, 2010.
- [10] LEE C, GE H Y, CHUNG E. What linguistic features distinguish and predict L2 writing quality? A study of examination scripts written by adolescent Chinese learners of English in Hong Kong[J]. *System*, 2021, 97: 102461.
- [11] 王海华, 周祥. 非英语专业大学生写作中词汇丰富性变化的历时研究[J]. *外语与外语教学*, 2012(2): 40 – 44.
- [12] 高霞. 不同水平学习者英语作文句法复杂度研究[J]. *外语教学与研究*, 2021, 53(2): 224 – 237; 319.
- [13] BULTÉ B, HOUSEN A. Conceptualizing and measuring short-term changes in L2 writing complexity[J]. *Journal of Second Language Writing*, 2014, 26: 42 – 65.

- [14] KYLE K, CROSSLEY S A. Automatically assessing lexical sophistication: indices, tools, findings, and application[J]. *TESOL Quarterly*, 2015, 49(4): 757 – 786.
- [15] ISHIKAWA S. Lexical development in L2 English learners' speeches and writings[J]. *Procedia-Social and Behavioral Sciences*, 2015, 198: 202 – 210.
- [16] 朱慧敏, 王俊菊. 英语写作的词汇丰富性发展特征——一项基于自建语料库的纵贯研究[J]. *外语界*, 2013(6): 77 – 86.
- [17] MCNAMARA D S, LOUWERSE M M, MCCARTHY P M, et al. Coh-Metrix: capturing linguistic features of cohesion[J]. *Discourse Processes*, 2010, 47(4): 292 – 330.
- [18] MARTÍNEZ A C L. Analysis of syntactic complexity in secondary education EFL writers at different proficiency levels[J]. *Assessing Writing*, 2018, 35: 1 – 11.
- [19] JIANG J Y, BI P, LIU H T. Syntactic complexity development in the writings of EFL learners: insights from a dependency syntactically-annotated corpus[J]. *Journal of Second Language Writing*, 2019, 46: 100666.
- [20] 鲍贵. 英语学习者作文句法复杂性变化研究[J]. *外语教学与研究*, 2009, 41(4): 291 – 297; 321.
- [21] LINNARUD M. *Lexis in Composition: A Performance Analysis of Swedish Learners' Written English*[M]. Malmö: Liber Forlag Malmö, 1986.
- [22] 雷蕾. 中国英语学习者学术写作句法复杂度研究[J]. *解放军外国语学院学报*, 2017, 40(5): 1 – 10; 159.
- [23] 文秋芳, 梁茂成, 晏小琴. *中国学生英语口笔语语料库 2.0 版* [M]. 北京: 外语教学与研究出版社, 2008.
- [24] LU X F. The relationship of lexical richness to the quality of ESL learners' oral narratives[J]. *The Modern Language Journal*, 2012, 96(2): 190 – 208.
- [25] COVINGTON M A, MCFALL J D. Cutting the Gordian knot: the moving-average type-token ratio (MATTR) [J]. *Journal of Quantitative Linguistics*, 2010, 17(2): 94 – 100.
- [26] MICHALKE M E, MIRISOLA A, BRULET A, et al. The koRpus package[EB/OL]. (2021-05-27) [2024-06-21]. <https://reaktanz.de/?c=hacking&s=koRpus>.
- [27] TAN L. NeoSCA (version 0.1.2) [CP/DK]. (2024-01-23) [2024-06-21]. <https://github.com/tanloong/neosca>.
- [28] WILCOX R R. *Introduction to Robust Estimation and Hypothesis Testing*[M]. 5th ed. San Diego: Elsevier, 2022.
- [29] COHEN J. *Statistical Power Analysis for the Behavioral Sciences*[M]. Hillsdale: Lawrence Erlbaum Associates, 1988.
- [30] LAUFER B. The lexical profile of second language writing: does it change over time?[J]. *RELC Journal*, 1994, 25(2): 21 – 33.
- [31] LEI S Y, YANG R Y. Lexical richness in research articles: corpus-based comparative study among advanced Chinese learners of English, English native beginner students and experts[J]. *Journal of English for Academic Purposes*, 2020, 47: 100894.
- [32] MANCILLA R L, POLAT N, AKCAY A O. An investigation of native and nonnative English speakers' levels of written syntactic complexity in asynchronous online discussions[J]. *Applied Linguistics*, 2017, 38 (1): 112 – 134.
- [33] 徐晓燕, 王维民, 熊燕宇, 等. 中国英语专业学生英语议论文句法复杂性研究[J]. *外语教学与研究*, 2013, 45(2): 264 – 275; 320.
- [34] 郑咏滢, 冯予力. 学习者句法与词汇复杂性发展的动态系统研究[J]. *现代外语*, 2017, 40(1): 57 – 68; 146.

(责编: 朱渭波)