

基于TAACO的文本连贯性与二语写作质量关系研究

赵琪源

中国矿业大学(北京)文法学院, 北京

收稿日期: 2022年6月16日; 录用日期: 2022年7月18日; 发布日期: 2022年7月27日

摘要

随着自然语言处理技术的发展, 文本连贯性的测量与研究已经成为二语写作评价的焦点。本研究以英语专业八级考试的写作文本为语料, 探究中国高水平英语学习者的文本连贯性与写作质量的关系。研究发现: 1) 局部连贯性与整体连贯性和写作质量呈正相关, 其中相关性最强的指标为Word2vec, 而篇章连贯性与二语写作质量不具有相关性; 2) 整体Word2vec、局部论元重叠、整体词元重叠等三项指标的联合可以解释写作成绩32.8%的方差。研究结论一定程度上可启示我国英语写作教学, 也可为作文自动评分工具的开发提供参考。

关键词

文本连贯性, 写作质量, TAACO

A Study on Relationship between Text Coherence and Second Language Writing Quality Based on TAACO

Qiyuan Zhao

School of Law and Humanities, China University of Mining & Technology (Beijing), Beijing

Received: Jun. 16th, 2022; accepted: Jul. 18th, 2022; published: Jul. 27th, 2022

Abstract

With the development of natural language processing technology, the measurement and study of text coherence has become the focus of second language writing evaluation. Taking TEM8 writings

as the research corpus, this study explores the relationship between text coherence and writing quality among high-level Chinese EFL learners. The results show that: 1) There is a positive correlation between local coherence and global coherence and writing quality, of which Word2vec is the most correlative indicator, while there is no correlation between text coherence and second language writing quality; 2) The combination of global Word2vec, local argument overlap and global lemmas overlap can explain 32.8% of the variance of writing performance. The conclusion of the study can enlighten the teaching of English writing in China to some extent, and also provide a reference for the development of automatic writing scoring tools.

Keywords

Text Coherence, Writing Quality, TAACO

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

文本连贯性是一种将文本中的思想有机联系在一起的逻辑关系,可以使文本在语义上具有连续性[1]。近些年,文本连贯性在二语写作领域,尤其是写作测试与评价方面已经成为热点问题。文本连贯性在读者理解文本含义以及作者的意图方面具有举足轻重的地位,因此研究其与写作质量之间的关系显得格外重要。

作为中国难度系数最高的英语等级考试,英语专业八级考试具有很高的权威性,其参与者为英语专业的大四学生,他们经过了大学四年专业的英语训练,能够代表中国高水平英语学习者。基于此,本研究将选取英语专业八级考试写作文本为语料研究中国高水平英语学习者的文本连贯性与写作质量的关系,以期对英语写作教学以及作文自动评分工具的开发提供一定启示。

2. 理论基础与文献综述

衔接与连贯最初由 Halliday 和 Hasan 提出,随即被广为研究。衔接与连贯是相互关联但又不同的结构,前者是指文本中的表层语言联系,而后者则是指文本在读者心中的心理表征质量,能够将文本的思想联系在一起[2]。一般来说,衔接往往与连贯密切相关,因为使用衔接手段可以表现出语篇的逻辑关系,便于读者理解。然而,与衔接手段不同的是,连贯是一个语义层面的概念,其一定意义上得助于衔接手段恰当地使用,但两者并非充分必要条件。

借助于自然语言技术对文本衔接特征的提取, Crossley 等人提出文本连贯性包括三个维度:局部连贯性、整体连贯性以及篇章连贯性[3]。局部连贯性指句子层面的局部衔接与连贯(例如,相邻句子间的词元重叠),整体连贯性则指段落层面的整体衔接与连贯(例如,相邻段落间的词元重叠),而篇章连贯性指整个篇章的衔接与连贯,它不仅仅局限在句子和段落层面而是立足于整个文本层面(例如,代词密度)。

目前国内外很多学者研究文本连贯性和二语写作质量(人工评分)的相关性,但研究发现各有不同。早期的研究主要集中于考察局部与篇章连贯性,大部分以 Halliday 和 Hasan 于 1976 年提出的五种衔接手段,即指称、连接、替代、省略和词汇衔接为理论基础[2]。这些研究通常考察文本中衔接手段使用的数量与二语写作质量的关系,其中大部分研究发现,局部与篇章连贯性与二语写作质量呈正相关。例如, Liu 和 Brane 以中国非英语专业本科生的 50 篇议论文为研究对象,考察他们的衔接手段使用情况及其与写作质量的关系,发现中国非英语专业本科生议论文的写作质量与衔接手段的使用量有显著的正相关性[4]。

Chanyoo 则以泰国英语专业本科生为研究对象, 考察了他们的学术论文的衔接特征与写作质量的关系, 同样发现学生的写作成绩与文本中使用的衔接手段总量呈显著的正相关[5]。可见, 这些研究表明无论是非学术语篇还是学术语篇, 文本连贯性与写作质量都存在着显著的正相关性。然而, 也有少部分研究发现文本中衔接手段的使用与文本质量不具有相关性。Alarcon 和 Morales 选取了菲律宾本科生 64 篇作文为语料, 发现衔接手段与学生作文质量没有显著的相关性[6], 该结论也与 Coskun 以及 Dastjerdi 和 Hayati 的研究发现一致。与 Alarcon 和 Morales 研究不同的是, Coskun 的研究语料则来自于土耳其和乌兹别克斯坦学生的英语写作[7], 而 Dastjerdi 和 Hayati 选取的研究对象则为伊朗非英语专业研究生[8]。

随着计算机技术的发展, 一些学者应用计算机技术开展研究, 把整体连贯特征纳入到他们的研究范围。这类研究考察的不再是衔接手段使用的数量, 而是一些基于工具所计算的连贯性各项指标的数值。一些研究发现文本连贯性与二语写作质量为正相关。例如, 杜慧颖和蔡金亭通过考察中国大学生英语议论文的语篇特征, 发现局部连贯性和整体连贯性都与写作质量存在显著的正相关[9]。梁茂成也以中国大学生为研究对象, 选取了 120 篇英语作文为研究语料, 但却发现高分作文整体连贯性较高, 而低分作文局部连贯性较高[1]。Crossley 和 McNamara 则以中国香港高中毕业的学生参加香港高级程度会考所写的英语作文为语料, 考察了文本连贯性对二语写作质量的预测效应, 发现一些局部和篇章连贯性指标, 如局部内容词重叠、局部积极逻辑连接词重叠、局部语义相似度等, 与香港高中生的英语写作质量呈显著的负相关[10]。

综上, 以往研究对文本连贯性与写作质量呈何种关系的结论不一, 且多数研究集中在研究局部和篇章连贯性特征, 鲜少研究整体连贯性特征, 更少以中国高水平英语学习者为研究对象。此外, 在文本连贯性指标对二语写作质量的贡献程度上也并未达成一致。因此, 本研究选取代表中国高水平英语学习者写作水平的英语专业八级考试作文文本为语料, 分析文本连贯性三个维度: 局部连贯、整体连贯、篇章连贯与写作质量的关系, 以期对写作教学以及作文评分工具的开发提供启示。

3. 研究设计

3.1. 研究问题与语料来源

本研究主要回答以下两个问题:

- 1) 中国高水平英语学习者文本连贯性和写作质量是否具有相关性? 若有, 具体体现在哪些指标上?
- 2) 中国高水平英语学习者文本连贯性对写作质量是否具有预测作用? 若有, 哪些指标对英语写作质量贡献率较大?

本研究语料选自《中国高校外语专业多语种语料库建设与研究——英语语料库》(以下简称 CEM 语料库)。CEM 语料库由英语专业四、八级考生作文以及英语专业八级翻译构成, 其中不仅包括文本, 还包括相应的性别、来源学校、分数等。作文的分数均由阅卷组专家评定, 因此具有很好的信度和效度。本研究随机选取某一年度的英语专业八级考生作文 200 篇, 建立 CEM_W8 子库。该年度考试题目要求作者以“On Ambition”为话题写一篇大约 400 字的议论文。CEM_W8 子库中考生平均分为 12.455, 最高分为 18, 最低分为 7, 标准差为 1.562, 分数呈正态性分布。

3.2. 文本连贯性指标选取

本研究文本连贯性分析工具采用 Crossley 等人开发的自然语言处理工具: 自动文本连贯性分析器 (Tools for the Automatic Analysis of Text Cohesion, TAACO), 版本为 2.0.4 [3]。TAACO 2.0.4 共涉及到 194 项指标, 分属于局部连贯特征、整体连贯特征以及篇章连贯特征三个维度。借鉴部寒和王立非[11]研究所选取的 11 项指标, 本研究去掉了与本研究所考察的文本连贯性关系不大的类符与形符比 (Type-Token

Ratio, TTR)指标, 加入语义相似度的 4 项指标, 最终选取 14 项连贯性指标进行测量, 具体如下。

局部连贯性指标选取局部词元重叠、局部论元重叠、局部同义名词重叠、局部同义动词重叠、局部 LSA、局部 Word2vec 等 6 项指标。与局部连贯特征选取指标一致, 整体连贯性指标也选取整体词元重叠、整体论元重叠、整体同义名词重叠、整体同义动词重叠、整体 LSA、整体 Word2vec 等 6 项指标。词元重叠是指相邻句子或段落间所有词元的重复。论元重叠指的是相邻句子或段落间论元的重复, 即名词与代词重叠。同义名词、同义动词重叠指的就是相邻句子或段落间的名词、动词的同义词的重复。本研究所有涉及到重叠的指标的分母均为句子数(sentence normed)或段落数(paragraph normed)。LSA 和 Word2vec 是两项基于自然语言处理技术的语义相似度指标, 前者的训练模型是基于奇异值分解(Singular Value Decomposition, SVD)的矩阵分解技术, 而后者则是基于神经网络对词的上下文训练从而得到词的向量化表示。

篇章连贯性则选取 2 项指标, 分别为连接词指标以及代词密度指标。连接词指标指的就是文本中所有具有连接意义的词的数量, 包括一些基本的连接词, 如“and”、“or”、“so”等, 也包括一些逻辑连接词, 如“alternatively”、“after all”、“despite the fact”等。代词密度指的是文本中第三人称代词数除以单词数。

3.3. 统计分析

通过 TAACO 2.0.4 计算语料库中各个写作文本的 14 项指标数值后, 随即采用 SPSS 28.0 进行统计分析。首先, 运行 Pearson 相关分析考察各项指标与写作质量的相关性。然后, 剔除不相关变量, 随即进行多元回归分析找出文本连贯性在多大程度上可以预测写作质量。最终得出相应的回归方程。

4. 结果

4.1. 文本连贯性和写作质量相关性分析结果

由表 1, 相关性分析结果显示, 文本连贯性 14 项指标中, 8 项指标与写作质量具有显著的正相关, 6 项指标与写作质量不具相关性, 没有指标与写作质量呈负相关。具体来看, 局部连贯性有 4 项指标与写作质量具有相关性, 由强到弱分别为局部论元重叠($r = 0.351, p < 0.01$)、局部同义名词重叠($r = 0.263, p < 0.01$)、局部 Word2vec ($r = 0.257, p < 0.01$)、局部词元重叠($r = 0.217, p < 0.01$), 而局部同义动词重叠、局部 LSA 与写作质量不具有相关性。与局部连贯性一致, 整体连贯性也有 4 项指标与写作质量呈正相关, 由强到弱分别为整体 Word2vec ($r = 0.491, p < 0.01$)、整体 LSA ($r = 0.312, p < 0.01$)、整体词元重叠($r = 0.199, p < 0.01$)、整体论元重叠($r = 0.160, p < 0.05$), 整体同义名词重叠与整体同义动词重叠与写作质量不具相关性。此外, 篇章连贯性的 2 项指标: 连接词和代词密度与写作质量不相关。综合来看, 与写作质量相关性强度排名前三的指标分别为整体 Word2vec、局部论元重叠、整体 LSA。

4.2. 文本连贯性和写作质量回归分析结果

如表 2 所示, 剔除不相关变量后, 进行线性回归分析, 共得出三个回归模型。模型 1 中包括整体 Word2vec 指标在内的一项文本连贯性指标, 在这种情况下 R^2 值为 0.237, 意味着整体 Word2vec 可以解释写作成绩 23.7% 的变异。模型 2 是由整体 Word2vec 和局部论元重叠构成, R^2 值为 0.28, 表明两者的联合可以解释 28% 成绩的变异。模型 3 中共包括整体 Word2vec、局部论元重叠、整体词元重叠等三项指标, 此时 R^2 值为 0.328, 是三个模型中的最大值, 意味着这 3 项指标的联合可以解释写作成绩 32.8% 的变异。

表 3 显示模型 3 通过 F 检验($F = 31.932, p = 0.000 < 0.05$), 说明该模型有效。回归模型中的三项预测变量(整体词元重叠、局部论元重叠、整体 Word2vec)均具有显著统计意义($t > 2$), 表明这三个指标对写作成绩产生显著的正向影响关系。模型中整体 Word2vec、局部论元重叠、整体词元重叠等三项指标的 VIF

值在 1 附近且小于 5, 意味着三项预测变量不存在共线问题。此外, 该模型的 D-W 值为 1.930, 在 2 附近, 说明模型不存在自相关性, 三项预测变量数据之间并不关联。模型 3 拟合程度较好, 解释了文本连贯性与写作质量的关系, 公式为: 写作分数 = $1.025 \times$ 局部论元重叠 + $0.083 \times$ 整体词元重叠 + $13.107 \times$ 整体 Word2vec - 1.569。

Table 1. Results of Pearson correlation analysis

表 1. 相关性分析结果

指标	平均值 \pm 标准差	相关系数	指标	平均值 \pm 标准差	相关系数
局部词元重叠	2.719 \pm 0.782	0.217**	整体词元重叠	17.003 \pm 3.807	0.199**
局部论元重叠	1.333 \pm 0.324	0.351**	整体论元重叠	4.562 \pm 1.322	0.160*
局部同义名词重叠	0.587 \pm 0.306	0.263**	整体同义名词重叠	6.911 \pm 3.265	0.120
局部同义动词重叠	0.528 \pm 0.299	0.031	整体同义动词重叠	10.747 \pm 6.204	-0.079
局部 LSA	0.667 \pm 0.040	0.108	整体 LSA	0.710 \pm 0.083	0.312**
局部 Word2vec	0.888 \pm 0.014	0.257**	整体 Word2vec	0.888 \pm 0.014	0.491**
连接词	0.087 \pm 0.018	0.115	代词密度	0.054 \pm 0.024	0.000

* $p < 0.05$; ** $p < 0.01$.

Table 2. Summary of multiple linear regression models

表 2. 多元线性回归模型汇总表

模型	R	ΔR^2	标准误	更改统计				
				R ² 更改	F 变化	df 1	df 2	Sig. F 更改
1	0.491 ^a	0.241	0.237	1.365	0.241	62.804	1	198
2	0.536 ^b	0.288	0.280	1.325	0.047	12.946	1	197
3	0.573 ^c	0.328	0.318	1.290	0.041	11.868	1	196

^a 预测变量: (常量), 整体 Word2vec; ^b 预测变量: (常量), 整体 Word2vec, 局部论元重叠; ^c 预测变量: (常量), 整体 Word2vec, 局部论元重叠, 整体词元重叠; 因变量: 写作分数。

Table 3. Coefficients of multiple linear regression

表 3. 多元线性回归系数表

	非标准化系数		标准化系数	t	p	VIF	R ²	ΔR^2	F
	B	标准误	Beta						
常数	-1.569	1.588		-0.988	0.324				
局部论元重叠	1.025	0.296	0.212	3.46	0.001	1.1	0.328	0.318	F(3, 196) = 31.932, p = 0.000
整体词元重叠	0.083	0.024	0.202	3.445	0.001	1.006			
整体 Word2vec	13.107	1.848	0.435	7.092	0.000	1.098			

因变量: 写作分数。

5. 讨论

本研究发现局部连贯性和整体连贯性与二语写作质量存在正相关性, 该结论部分验证了杜慧颖和蔡金亨[9]的结论, 进一步证实了局部连贯性与整体连贯性对写作质量具有一定的影响。然而, 基于研究工具使用的不同, 本研究所发现的与二语写作质量呈正相关的指标与他们的研究发现不太一致。本研究基于 TAACO 工具发现局部连贯性指标中的局部论元重叠、局部同义名词重叠、局部 Word2vec、局部词元重叠以及整体连贯性指标中的整体 Word2vec、整体 LSA、整体词元重叠、整体论元重叠与写作质量具有正相关性, 而他们则基于 Coh-Metrix 工具发现局部 LSA、所有句子间 LSA、论元重叠相邻句子比例、论元重叠句子比例、词干重叠相邻句子比例、词干重叠句子比例以及整体 LSA 指标与写作质量正相关。因此, 本研究丰富了文本连贯性与二语写作质量的相关性指标。然而, 两项研究还存在一些分歧, 表现在局部 LSA 上。可以看出, 在相同的计算指标中, 本研究发现局部 LSA 和写作质量不具有相关性, 而他们则发现局部 LSA 与写作质量呈正相关。因此, 局部 LSA 是否与写作质量存在相关性还需要在接下来的研究中进一步验证。

本研究相关性分析结果与 Crossley 和 McNamara [10]的研究发现相悖。他们发现中国香港高中生的局部和篇章连贯性指标与写作质量呈负相关, 而本研究则发现中国高水平学习者的文本连贯性与写作质量呈正相关。研究对象选取的不同可能是造成这两个研究结论不一致的原因。因此, 本研究认为不同水平写作的考察侧重点可能有所差异, 所以现有的作文自动评分工具可能需要进一步优化, 将不同学习阶段的测评指标纳入评分系统中, 以便对具体阶段水平的作文具体分析, 而不是仅仅应用一个模型对其进行评分。此外, 相关性分析结果表明所有连接词以及代词密度两项篇章连贯性指标与二语写作质量不具有相关性。该发现与相关研究发现一致[3], 进一步说明了文本中连接词的运用以及代词密度不会影响文章的质量。

在回归模型中, 我们可以发现整体 Word2vec、局部论元重叠、整体词元重叠等三项指标的联合可以解释写作成绩的 32.8%变化原因, 说明局部连贯性与整体连贯性的联合更能影响写作质量。因此, 学习者要在写作中兼顾局部连贯性与整体连贯性。此外, 自动评分工具的开发可以引入这三项文本连贯性指标, 使机器评分和人工评分更为接近。

6. 结语

本研究以英语专业八级考试写作文本为语料, 探究了中国高水平学习者文本连贯性与写作质量的关系, 发现: 1) 局部连贯性和整体连贯性各有四项指标与写作质量呈正相关关系, 局部连贯指标分别为局部论元重叠、局部同义名词重叠、局部 Word2vec、局部词元重叠, 整体连贯性指标分别为整体 Word2vec、整体 LSA、整体词元重叠、整体论元重叠。综合来看, 相关性强度排名前三的指标分别为整体 Word2vec、局部论元重叠、整体 LSA; 2) 多元回归分析共产生三个回归模型, 其中整体 Word2vec、局部论元重叠、整体词元重叠等三项指标的联合可以解释写作成绩 32.8%的变异。标准化回归方程为: 写作分数 = $1.025 \times$ 局部论元重叠 + $0.083 \times$ 整体词元重叠 + $13.107 \times$ 整体 Word2vec - 1.569。

基于研究发现, 本研究为写作教学与自动评分工具的开发提供几点启示。首先, 教师在写作教学中应提醒学习者不要仅仅关注局部的连贯, 更要关注整体的连贯。其次, 自动工具的开发不能仅仅应用一个单独的模型进行评分, 应使不同学段评分规则差异化, 提供可选择评分模型的功能。最后, 可以将整体 Word2vec、局部论元重叠、整体词元重叠三项指标联合的回归模型引入到中国高水平学习者写作自动评分工具的开发中。

虽然本研究探究了中国高水平英语学习者的文本连贯性与写作质量之间的关系, 但鉴于所选取的语

料库的特点, 本研究仅仅考察了文本连贯性与总体写作成绩的关系, 并未考虑到篇章结构、内容、语言等单项分数。为此, 接下来的研究应细化写作评分, 将连贯性与写作质量各方面的评分分别进行相关性分析, 找出连贯性具体对哪一单项的评分会产生影响, 以便更精细地指导教学以及自动评分工具的开发。

参考文献

- [1] 梁茂成. 学习者书面语篇连贯性的研究[J]. 现代外语, 2006, 29(3): 284-292.
- [2] Halliday, M.A.K. and Hasan, R. (1976) *Cohesion in English*. Longman, London.
- [3] Crossley, S.A., Kyle, K. and McNamara, D.S. (2016) The Tool for the Automatic Analysis of Text Cohesion (TAACO): Automatic Assessment of Local, Global, and Text Cohesion. *Behavior Research Methods*, **48**, 1227-1237. <https://doi.org/10.3758/s13428-015-0651-7>
- [4] Liu, M. and Braine, G. (2005) Cohesive Features in Argumentative Writing Produced by Chinese Undergraduates. *System*, **33**, 623-636. <https://doi.org/10.1016/j.system.2005.02.002>
- [5] Chanyoo, N. (2018) Cohesive Devices and Academic Writing Quality of Thai Undergraduate Students. *Journal of Language Teaching and Research*, **9**, 994-1001. <https://doi.org/10.17507/jltr.0905.13>
- [6] Alarcon, J.B. and Morales, K.N.S. (2011) Grammatical Cohesion in Students' Argumentative Essay. *Journal of English and Literature*, **2**, 114-127.
- [7] Coskun, E. (2011) Cohesion in Compositions of Turkish and Immigrant Students. *Educational Sciences: Theory and Practice*, **11**, 892-899.
- [8] Dastjerdi, H.V. and Hayati, S.S. (2011) Quality of Iranian EFL Learners' Argumentative Essays: Cohesive Devices in Focus. *Mediterranean Journal of Social Sciences*, **2**, 65.
- [9] 杜慧颖, 蔡金亭. 基于 Coh-Metrix 的中国英语学习者议论文写作质量预测模型研究[J]. 现代外语, 2013, 36(3): 293-300, 331.
- [10] Crossley, S.A. and McNamara, D.S. (2012) Predicting Second Language Writing Proficiency: The Roles of Cohesion and Linguistic Sophistication. *Journal of Research in Reading*, **35**, 115-135. <https://doi.org/10.1111/j.1467-9817.2010.01449.x>
- [11] 部寒, 王立非. 基于语料库的中美企业财务语篇可读性对比分析[J]. 解放军外国语学院学报, 2021, 44(1): 71-78, 128.