

基于Coh-Metrix的中国英语学习者学术论文引言部分衔接手段的研究

李佳琪, 马建军, 朱穆朗玛

大连理工大学, 辽宁 大连

收稿日期: 2022年11月8日; 录用日期: 2022年12月6日; 发布日期: 2022年12月15日

摘要

衔接手段的运用一直是学术论文写作的难点。本文基于Coh-Metrix分析了中国英语学习者(L2)学术论文引言部分的衔接手段。研究将50篇理工科博士生的引言和英语母语者(L1)引言进行比较, 基于30个Coh-Metrix指标, 从语法衔接、词汇衔接、语义衔接和情景模型衔接四个维度解析了引言的显性衔接和隐性衔接手段。研究发现, L2在引言衔接手段上与L1有显著差异, 存在使用过度或使用不足等问题。显性衔接问题主要体现在过度使用连词、所指总数及第一人称代词复数使用不足、用词过度重复; 隐性衔接问题主要在于句间语义相似度过高、旧信息占比过高、过度使用因果虚词、时态一致性不足。研究表明, 英语水平、写作教学、写作意图以及汉语负迁移等是影响L2使用衔接手段的主要因素。

关键词

衔接手段, 显性衔接, 隐性衔接, 学术论文引言, Coh-Metrix

A Coh-Metrix Study on Cohesive Devices in Chinese English Learners' Research Article Introductions

Jiaqi Li, Jianjun Ma, Zhumulangma

Dalian University of Technology, Dalian Liaoning

Received: Nov. 8th, 2022; accepted: Dec. 6th, 2022; published: Dec. 15th, 2022

Abstract

The use of cohesive devices has always been a difficulty in research article writing. Based on

文章引用: 李佳琪, 马建军, 朱穆朗玛. 基于 Coh-Metrix 的中国英语学习者学术论文引言部分衔接手段的研究[J]. 现代语言学, 2022, 10(12): 2989-3000. DOI: 10.12677/ml.2022.1012403

Coh-Metrix, this paper analyzes the cohesive devices used by Chinese EFL learners (L2) in the introductions of their research articles. This study compares 50 introductions of PhD students of science and engineering with those of native English speakers (L1), and based on 30 Coh-Metrix indices, it analyzes the explicit and implicit cohesive devices from four dimensions: grammatical cohesion, lexical cohesion, semantic cohesion and situation model cohesion. Results show that L2 introductions are significantly different from L1 introductions in terms of the use of cohesive devices, indicating problems of both overuse and underuse. The main problems of explicit cohesion lie in the overuse of connectives, insufficient use of references and first-person plural pronouns, and excessive lexical repetition, while the main problems of implicit cohesion lie in high semantic similarity between sentences, high given-new information ratio, overuse of causal particles, and low temporality. These findings suggest that factors such as the English level, teaching of writing, writing intention and negative transfer of Chinese may affect how L2 use cohesive devices.

Keywords

Cohesive Devices, Explicit Cohesion, Implicit Cohesion, Research Article Introduction, Coh-Metrix

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

衔接是一种存在于语言中的机制，对语篇结构和语篇质量起着重要作用[1]。学者们对衔接手段的分类方式较为多元，其中占主流的观点主要有以下三类：Halliday 和 Hasan 将衔接分为语法衔接和词汇衔接，胡壮麟[2]认为语篇衔接和连贯是多层次的，并将其细分为社会符号层、语义层等，而张德禄和刘汝山[3]则将衔接手段分为显性衔接和隐性衔接。

事实上，在写作中实现衔接手段的灵活运用并非易事，尤其是对于有学术论文写作需求的英语学习者而言。有调查表明，英语学习者(L2)与母语者(L1)在衔接手段的使用上存在显著差异：其中 L2 在英语写作中对于上位词、论元重复、因果动词等衔接手段的使用频次显著低于 L1 [4]；在所指的使用频次方面也表露出明显的使用偏好，L2 使用“we”的总数多于 L1 [5] [6]，而使用的所指总数显著少于 L1 [7]。因此，时下对于英语学习者学术论文中衔接手段使用情况的探究依旧不容忽视。

然而通过梳理相关文献可知，首先，就研究语料而言，既有研究大多以学术论文全篇[8]、摘要[9]或研究方法部分[10]作为分析文本，对学术论文引言部分(以下简称引言)衔接手段的研究则相对不多。引言是学术论文的开篇布局，其写作质量的高低直接决定了论文的可读性和逻辑性[11]，而衔接手段则在表现引言的连贯性与逻辑性过程中发挥着不可替代的作用[12]，因此，很有必要探索英语学习者进行学术论文引言部分写作时是否充分使用了衔接手段。

其次，就研究内容而言，既往针对学术写作衔接手段的研究多围绕显性衔接手段，且研究对象较为单一，主要集中在所指和连词等衔接手段，并聚焦于写作立场或身份构建研究，对于其他衔接手段，尤其是隐性衔接手段的探索则相对不多。因此，本研究将不仅考察中国英语学习者在学术写作时对于显性衔接的使用，其对隐性衔接的运用也是我们意欲探索的重点之一，以期对这一群体衔接手段的应用能力做出相对全面的诊断。

再次，就研究方法而言，过去许多研究或依赖人工标注文本的语言特征，或在文本之间进行直观的比较[13]。近年来，随着科技和学科间的融合发展，Coh-Metrix 逐渐成为国内外学者研究衔接手段时较为

青睐的工具，故本研究也选取了 Coh-Metrix 作为研究工具，对 L2 的引言衔接手段使用情况加以探究。

因此，本文以中国某高校 50 位理工科博士研究生为研究对象，自建引言学习者语料库和对比语料库，从显性衔接和隐性衔接两个角度，使用在线文本分析工具 Coh-Metrix 探究 L2 引言写作中衔接手段的使用特征，以期发现存在的问题及其影响因素。具体来看，本文的研究问题包括：

- 1) L2 和 L1 在引言中使用的衔接手段有哪些特点？
- 2) L2 与 L1 在引言中使用衔接手段的偏好上是否存在显著差异？
- 3) 影响 L2 在引言中使用衔接手段的因素是什么？

2. 衔接手段

Halliday 和 Hasan 将衔接定义为语篇中某一要素的解释依赖于另一要素的解释而产生的意义关系，该理论为衔接手段提供了有效的分析框架，不过仍存在一些待完善之处，如：未完全形成系统的分析框架、分析的具体化程度不够等。有鉴于此，张德禄和刘汝山进一步扩大了衔接手段的范围，将衔接手段分为显性衔接和隐性衔接。其中，显性衔接指的是由两个或两个以上的项目形成的衔接纽带，包括语法衔接、词汇衔接、音系衔接；而隐性衔接是指由意义空缺形成的衔接纽带，包括情景性省略和背景性省略。比较而言，后者在衔接手段的分类中增加了语音和意义两个层面，在内容上更加完善，但值得注意的是，其对隐性衔接的衔接手段的界定略显主观，且在操作性方面也存在一定的局限性。因此，本文在借鉴以上衔接理论框架的基础上，结合 Coh-Metrix 的指标，重新定义了隐性衔接手段。其中，省略，替代和搭配三种显性衔接手段只能进行人工分析，因此本研究暂不考虑，故本文在显性衔接方面主要侧重于考察连词、所指和重复，而隐性衔接方面则包括语义衔接和情景模型衔接(见图 1)。

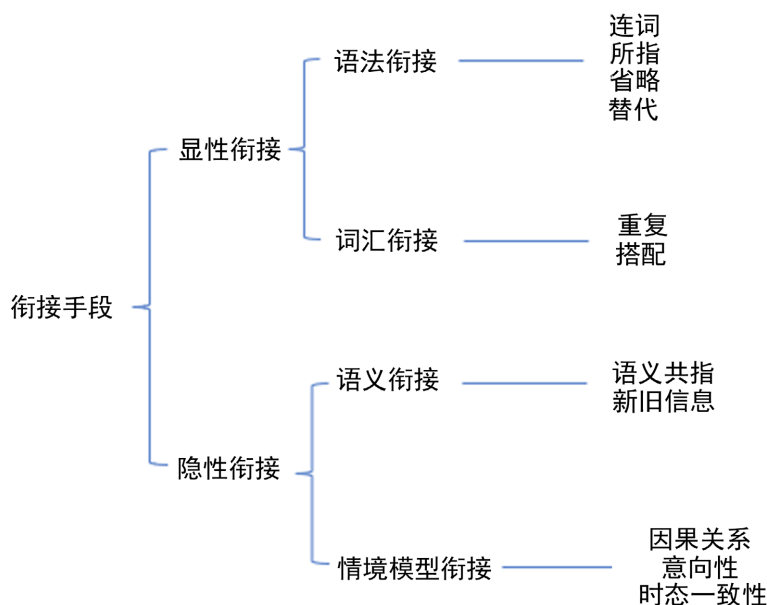


Figure 1. Cohesive devices for this study

图 1. 本研究的衔接手段

3. 研究方法

3.1. 研究工具

本研究借助 Coh-Metrix 3.0 对语料进行分析。Coh-Metrix 3.0 是衔接研究中一种被广泛使用的在线文

本分析工具, 它可以提供 106 个指标, 涵盖词汇、句子、段落、整体语篇的各项衔接分析。此外, 本研究使用 SPSS 25.0 作为数据分析工具, 进行独立样本 T 检验来判断 L2 与 L1 在引言部分使用衔接手段偏好上的显著差异。

3.2. 研究语料

本研究基于 L2 引言语料库和 L1 引言对比语料库两个自建语料库探讨研究问题。语料库的语料主要涉及化工和能源学科。L2 语料库的语料来自某高校博士生学术英语课程学生提交的期末论文引言部分, 共 50 篇, 涉及化工与环境生命、材料科学与工程、环境工程和能源与动力四个专业(详见表 1)。L1 语料库的语料选取于 2017~2021 年间发表在国际顶级英文期刊《自然化学》(*Nature Chemistry*)与《自然能源》(*Nature Energy*)上的学术论文, 每年各随机选取 5 篇引言, 共 50 篇。期刊的选择条件是: 1) 在美国或英国出版; 2) 在化工与能源学科方面享有重要声誉并具备高影响因子的国际顶级期刊。

语料在输入 Coh-Metrix 分析之前, 转为 txt 格式, 并去除了两个语料库中的多余字符, 如: 表格、公式、图片等, 使用分段符为段落标记, 语料库信息详见表 1。

Table 1. L2 introduction corpus & L1 introduction comparative corpus

表 1. L2 引言语料库 & L1 引言对比语料库

	专业/期刊	篇数					合计(数)	总词数(词)
		2021	2020	2019	2018	2017		
L2 引言语料库	化工与环境生命	28	/	/	/	/	50	25809
	材料科学与工程	9	/	/	/	/		
	环境工程	6	/	/	/	/		
	能源与动力	7	/	/	/	/		
L1 引言对比语料库	Nature Chemistry	5	5	5	5	5	50	28265
	Nature Energy	5	5	5	5	5		

表 2 表明了 Coh-Metrix 对语料库的描述性分析, 前四个指标可以看出, L2 引言语料库与 L1 引言对比语料库中的各项为参照, 均无显著差异($ps > 0.05$)。表 2 中后三个指标考察词汇多样性, 可以看出, 所有字符在文本中保持给定类型标记比率值的连续字串的平均长度(LDMTLD)和所有字符的类型标记比率随机样本与理想类型标记比率曲线拟合值(LDVOCD)均存在显著性差异($ps < 0.01$), 说明 L2 词汇多样性弱于 L1。因此, 如果 L2 在引言中与 L1 衔接手段存在显著差异, 词汇多样性可能是原因之一。

Table 2. Descriptive analysis of corpus

表 2. 语料库描述性分析

	文本数量	DESWC	DESSC	DESPC	DESSL	LDTTRa	LDMTLD	LDVOCD
L2 引言语料库	50	516.18	19.96	3.22	25.46	0.51	92.94	94.45
L1 引言对比语料库	50	565.30	23.18	3.72	24.63	0.52	113.75	116.24
p 值	/	0.087	0.054	0.102	0.243	0.602	0.000	0.000

3.3. 数据分析

语料净化后, 我们将其输入 Coh-Metrix 以备分析。研究共涉及 Coh-Metrix 30 个指标, 包括五大类: 连词、所指、重复、潜在语义分析和情境模型, 具体见表 3。之后使用 SPSS 25.0 对数据进行独立样本 T 检验, 分析 L2 和 L1 在衔接手段使用上是否存在显著差异。

Table 3. The selected Coh-Metrix cohesive devices for this study

表 3. 本研究选取的 Coh-Metrix 指标

类别	数目	Coh-Metrix 指标
连词 (connective)	6	连词总数(CNCALL)、因果连词(CNCCaus)、逻辑连词(CNCLogic)、转折及对比连词(CNCADC)、时间连词(CNCTemp)、递进连词(CNCAdd)
所指 (reference)	6	所指总数(WRDPRO)、第一人称单数(WRDPRP1s)、第一人称复数(WRDPRP1p)、第二人称(WRDPRP2)、第三人称单数(WRDPRP3s)、第三人称复数(WRDPRP3p)
重复 (overlap)	8	名词重复句子比例(CRFNOa)、名词重复相邻句子比例(CRFNO1)、论元重复句子比例(CRFAOa)、论元重复相邻句子比例(CRFAO1)、词干重复句子比例(CRFSOa)、词干重复相邻句子比例(CRFSO1)、实词重复句子比例(CRFCWOa)、实词重复相邻句子比例(CRFCWO1)
潜在语义分析 (latent semantic analysis)	4	相邻句子语义相似度(LSALL1)、段落内所有句子语义相似度(LSASSp)、相邻段落语义相似度(LSAPP1)、旧信息占比(LSAGN)
情境模型 (situation model)	6	因果动词(SMCAUSv)、因果动词及因果虚词(SMCAUSvp)、意向动词(SMINTEp)、因果关系(SMCAUSr)、意向性(SMINTEr)、时态一致性(SMTEMP)

4. 结果与讨论

4.1. 衔接手段总体特征

因为各类指标计算方式不同, 表 4 选取了其中五个高度具备代表性的衔接手段指标衡量总体特征, 其中, 总数是标准化频数(每千词的频数), 因果关系比例指因果虚词和因果动词之比。表 4 显示, L2 运用的衔接手段种类丰富, 已经具备在引言中使用显性衔接和隐性衔接的能力, 以及促成引言连贯的意识。其次, L2 和 L1 在各类衔接手段上都具有显著性差异($ps < 0.05$), 表明 L1 存在连词使用过度、所指使用不足、实词重复过度、语义重复过度以及因果虚词使用过度等问题。以下进一步探讨 L2 与 L1 在引言使用各类衔接手段的特征及差异。

Table 4. Overall characteristics of cohesive devices

表 4. 衔接手段总体特征

	显性衔接			隐性衔接	
	连词总数	所指总数	实词重复句子比例	段落内所有句子语义相似度	因果关系比例
L2 引言语料库	89.86	10.35	0.09	0.28	0.60
L1 引言对比语料库	83.28	14.86	0.07	0.24	0.38
p 值	0.049	0.000	0.000	0.001	0.001

4.2. 显性衔接特征

4.2.1. 连词

连词在增强句子之间的凝聚性及提供文本组织线索两个方面起重要作用[14]。Coh-Metrix 计算其总数及五类连词(因果连词、逻辑连词、转折及对比连词、时间连词、递进连词)的每千字频数。值得一提的是, 这些类别的连词或有重复, 如 *and* 即可以是逻辑连词, 也可以是递进连词。

研究显示(表 5), L2 使用的连词类型十分丰富, 说明 L2 意识到了运用连词实现衔接的重要性, 其中逻辑连词使用最多。独立样本 T 检验结果表明, L2 在因果连词、逻辑连词、转折及对比连词、时间连词和递进连词五项上与 L1 使用的数量并无显著差异($p > 0.05$), 然而, 二者在连词总数方面有显著性差异($p = 0.049 < 0.05$), 表明 L2 连词使用总数高于 L1。

Table 5. Connective word frequency statistics (Per Thousand Words)

表 5. 连词词频统计(每千词)

连词(平均值)	连词总数	因果连词	逻辑连词	转折及对比连词	时间连词	递进连词
L2 引言语料库	89.86	28.66	32.77	11.75	14.1	49.81
L1 引言对比语料库	83.28	25.76	34.41	14.72	12.39	49
p 值	0.049	0.105	0.424	0.056	0.179	0.737

对语料的进一步分析可知, 我国英语学习者在引言写作中确实存在连词过度使用现象, 且经常会因为使用过多, 反而致使文章连贯性较差。如例 1 中 *and* 多余, 只用 *moreover* 表示递进关系即可; 例 2 用 *although* 表示转折及对比关系即可, *however* 多余。

例 1 (L2): And moreover, g-C3N4 has an appropriate bandgap energy (2.7 eV) that can effectively absorb visible light.

例 2 (L2): Although it is widely used in all walks of life for the Gas metal arc welding (GMAW). However, the demand for efficient welding methods is becoming more and more urgent with the development of society.

4.2.2. 所指

所指是指用代词作为语法手段来表示语义关系。其中, 人称代词是所指中的一类, 具体分为第一人称单数(“I”, “me”), 第一人称复数(“we”, “us”), 第二人称(“you”), 第三人称单数(“he”, “she”, “it”), 第三人称复数(“they”, “those”)。Coh-Metrix 计算所指总数及每类人称代词的每千字频数(表 6)。

表 6 表明, L2 和 L1 所指总数及第一人称复数两个指标存在显著差异($p < 0.01$), 也即 L2 的所指总数及第一人称复数 “we” 和 “us” 使用不足。另外值得注意的是, 二者均未使用第二人称代词, 而其它三个指标独立样本 T 检验的结果无显著差异($p > 0.05$)。

Table 6. Reference word frequency statistics (Per Thousand Words)

表 6. 所指词频统计(每千词)

所指(平均值)	所指总数	第一人称单数	第一人称复数	第二人称	第三人称单数	第三人称复数
L2 引言语料库	10.35	0.1	1.89	0	0.27	2.53
L1 引言对比语料库	14.86	0.2	8.28	0	0.02	3.42
p 值	0.000	0.394	0.000	/	0.157	0.166

研究发现, L2 多采取被动语态等方式, 尽量减少“we”的使用。例 3-6 表明, L2 应用了相当多的被动句, 让“endeavors, technology”等做主语, 以尽量减少“we”做主语的表达。即使在使用“we”时, 也借助了“as we all know”以削弱主观性, 增强认同; 或者利用“in the previous paper”以增强表达的严谨性。

例 3 (L2): Great endeavors have been devoted to the design and synthesis of novel materials that help to convert solar energy into chemical energy.

例 4 (L2): Optical imaging probe and technology are expected to achieve early cancer diagnosis and image-guided treatment.

例 5 (L2): As we all know, compared with amorphous glass, crystalline phase with the same chemical composition has smaller dielectric loss.

例 6 (L2): In the previous paper we reported a phase change material containing carbon-carbon double bond, which could retain shape after polymerization.

反观 L1, 直接在引言中应用“we”, 表明主观能动性, 强调大家作为一个团队共同研究。如例 7 所示, L1 甚至在一个段落中接连使用“we”做主语, 直观清楚地表明团队的研究步骤。

例 7 (L1): We first review the development of standardized qualification tests for commercial PV module reliability, including their shortcomings and failures. We show how stress tests can reveal material, cell and module degradation mechanisms and can be used to extract module lifetimes. We then discuss published stability studies of pre-commercial PSCs, and highlight key stress factors and the associated degradation mechanisms.

这一结果与刘应亮和陈洋的研究结果一致, 却与王晶晶和姜峰的结果相反。王晶晶和姜峰认为理工科博士 L2 使用指称的数量与其身份建构和写作立场息息相关, L2 在写作时会多使用“we”刻意回避作者身份并弱化自己的贡献, 是对自身学术立场不够自信的表现。刘应亮和陈洋则认为这与传统的儒家文化有关, L2 受其影响倾向于用团体来指代自己, 避免流露个人感情。而我们认为这与两个因素有关。第一, 写作教学。学术语篇的客观、理性的特点决定了传统的英语学术写作教学往往建议学生在文内以被动语态代替第一人称, 这就使得学习者在学术写作时存在过度规避使用第一人称的问题。正如王月丽和徐宏亮[15]所指出的那样: L2 在学术写作中使用第一人称代词的障碍很大程度上是由于传统学术写作的负面影响。第二, 写作意图。L1 在引言中使用较多的第一人称复数意在缩短读者和作者之间的距离。Harwood [16]和 Wales [17]认为, 使用“we”能够帮助作者实现一些具体的目的, 如让读者有参与研究的感觉, 缓解作者在学术信心不足时选择“1”所带来的心理压力等。

4.2.3. 重复

重复即为词汇的重复。Coh-Metrix 八个词汇衔接指标分为两个维度: 局部重复和整体重复, 局部重复指的是相邻句子名词、论元、词干、实词重复的比例, 而整体重复是指名词、论元、词干、实词重复句子比例(表 7)。

由表 7 可以看出, L2 和 L1 在八个指标均具有显著性差异($ps < 0.01$)。无论是局部重复还是整体重复, L2 引言中的名词重复、论元重复、词干重复和实词重复均高于 L1。这表明 L2 的引言词汇多样性不够丰富且重复用词较多, 存在过度重复的问题, 该结果也与表 2 结果一致。

这一结果与许双如[18]的研究结果一致, 他也提出 L2 使用了太多简单容易的表达来实现重复。如例 8 表明 L2 反复使用 research, basic 等非术语实词, 造成了名词和实词过度重复的问题。

例 8 (L2): In the future, we believe that NIR- II imaging will contribute to a wide range of basic and clinical

research. In oncology research, NIR- II imaging provides the possibility of three-dimensional structure, vascular distribution, blood flow and dynamic immune cell infiltration process of tumor; in regenerative medicine, non-invasive NIR- II imaging will also play an important role in exploring the development process of embryo and organ, stem cell lineage and other basic biological problems; NIR- II imaging also has great significance in neuroscience research it has broad application prospects.

Table 7. Reiteration ratio statistics

表 7. 重复比例统计

重复 (平均值)	名词重复 相邻句子 比例	论元重复 相邻句子 比例	词干重复 相邻句子 比例	实词重复 相邻句子 比例	名词重复 句子比例	论元重复 句子比例	词干重复 句子比例	实词重复 句子比例
L2 引言语 料库	0.67	0.72	0.79	0.13	0.54	0.58	0.67	0.09
L1 引言对 比语料库	0.52	0.58	0.67	0.09	0.42	0.48	0.57	0.07
p 值	0.000	0.000	0.000	0.000	0.001	0.003	0.002	0.000

而 L1 通过灵活的句子主语调整, 使用同/近义词替换等方法巧妙避免了过度重复, 使其表达更加灵活, 语言可读性更强。如例 9 中, L1 使用动名词短语“bringing them to market”做主语巧妙地概括上一句的内容, 自然地与下文衔接。同时, 为避免反复使用同一词汇, 使用同/近义词进行转换, 如: 后文提到的“wearable”与前文的“high performance-to-weight ratio”相呼应, “portable”呼应“mechanical flexibility”等。

例 9 (L1): Organic photovoltaics (OPVs) are a class of solar cells being developed for applications that require high performance-to-weight ratio, mechanical flexibility and/or semi-transparency with low production costs. Bringing them to market could provide a versatile clean energy solution that seamlessly integrates into buildings and windows, while also enabling new wearable, portable electronics and internet-of-things devices.

致使过度重复问题出现的一个最主要的原因在于汉语负迁移。汉语作为分析性语言, 多使用重复实现衔接, 而英语是曲折性语言, 多通过灵活调整句子结构来实现衔接[19]。连淑能[20]也认为, 重复是汉语的一个特点, 有助于达到特殊的效果。长期以来, 虽然 L2 意识到英汉两种语言的不同, 但是缺乏英语作为母语的学习环境, 汉语意合且常使用重复的特点造成了语言的负迁移, 使得 L2 在写作时受到了母语的较大影响, 无意识地倾向词汇重复。

4.3. 隐性衔接特征

4.3.1. 语义衔接

本研究从语义共指和新旧信息两方面评价语义衔接(LSA)。前者涉及三个指标: 相邻句子语义相似度、段落内所有句子语义相似度和相邻段落语义相似度, 可以对语义上相似但在词形可能没有关联的词进行追踪; 后者则通过旧信息占比计算每个句子中的旧信息比例。

表 8 表明, L2 语义共指中有两个指标与 L1 有显著区别($ps < 0.01$), 相邻段落语义相似度($p = 0.642 > 0.05$)则与 L1 无区别, 说明 L2 引言中同一段落中相邻句子的语义相似度比 L1 更密切, 引言中的句间语义相似度过高, 而相邻段落语义相似度则处理得当。

Table 8. Semantic cohesion ratio statistics
表 8. 语义衔接比例统计

语义衔接(平均值)	相邻句子语义相似度	段落内所有句子语义相似度	相邻段落语义相似度	旧信息占比
L2 引言语料库	0.32	0.28	0.57	0.35
L1 引言对比语料库	0.26	0.24	0.58	0.33
p 值	0.000	0.001	0.642	0.039

比如, 在例 10 中有三句, L2 在第一句第一次提到 “important source of energy” 时是新信息, 第二句紧接着提到 “important role as energy carrier”, 第三句又重复提到 “an important raw material”, “source”、“carrier” 和 “raw material” 这三个表达语义重复, 高度关联。同理, 第二句中 “environmentally friendly energy” 与第一句中的 “low environmental pollution” 语义高度关联。这些语义重复造成 L2 相邻句子相似度或段内句子相似度过高, 同时也使 L2 引言更加容易理解。

例 10 (L2): Hydrogen is considered a very important source of energy with a high potential of energy generation, very low environmental pollution, and large quantity all around the world. European Commission foresee that hydrogen will play an increasingly important role as energy carrier, providing environmentally friendly energy to end-users. Moreover, hydrogen is an important raw material in oil, gas and petrochemical as well as other industries.

在新旧信息方面, 从旧信息占比指标可以看出(表 8), L2 的旧信息占比高于 L1, 二者有显著差异($p = 0.039 < 0.05$)。旧信息占比越高, 该指标的比例越接近 1, 意味着文本中的已知信息越多, 因此就会更加连贯。

这一差异凸显 L2 和 L1 的英语水平差异。例 11 表明, L2 在一个段落中四句话的主语均与 MOF 相关, 仅仅依靠机械性地重复来实现句间衔接, 这使得 L2 的引言更加连贯, 更容易理解, 但隐性衔接不足, 表达较为浅显。

例 11 (L2): Metal-organic frameworks (MOFs) have received much attention in recent years due to their applications in gas storage, separations, catalysis, molecular sensing, and drug delivery. Metal-organic frameworks (MOFs) are hybrid materials that consist of secondary building units (SBUs) and organic linkers. The rational design of MOFs with tunable properties through a selective combination of metal ions and organic ligands has produced materials useful for various applications. However, the development of MOF photocatalysts for photoredox and nickel dual-catalyzed reactions is still in its infancy.

然而, 在例 12 中 L1 通过调换主语, 运用 “theoretical studies、computational studies、such studies、theoretical efforts” 总结前句已知信息, 从而突出新信息, 让已知信息和新信息更加分明, 句间语义衔接处理得当, 提到的新信息多, 使旧信息占比也较低。

例 12(L1): In spite of major progress in synthesizing and discovering catalysts for electrochemical processes, the interpretation of experiments on these complex materials and environments remains challenging. Moreover, design rules for new materials and molecules are often difficult to derive based only on experimental data. Theoretical and computational studies are playing an increasingly important role in interpreting and complementing experiments in the field of heterogeneous catalysis. Such studies are also providing insights that are leading to the formulation of catalyst design rules. The combination of experimental and theoretical efforts enhances the discovery process and technological innovation.

4.3.2. 情景模型衔接

情景模型是认知科学家用来评价文本的心理表征的术语。Coh-Metrix 的情景模型衔接能够超越明确的文本进入推理的领域。当一个或多个背景模型维度出现不连续时,就会出现凝聚力或连贯性的中断。每当这种不连续性发生时,重要的是要出现连接词、过渡性短语、副词等,向读者传达存在不连续的信息。Coh-Metrix 提供三个指标:因果关系、意向性、时态一致性来评测文本的情景模型衔接。因果关系指标是指因果虚词和因果动词之比,意向性指标计算的是意向性虚词与意向性动词之比,时态一致性指标考察时态的重复。

表 9 表明 L2 的因果动词及因果虚词总和、因果关系和时态一致性三个指标与 L1 有显著差异($p < 0.05$)。L2 因果动词及因果虚词总和以及因果关系比例比 L1 高,而二者因果动词使用数量相当($p = 0.474 > 0.05$),说明 L2 在引言中使用的因果虚词数量显著多于 L1,从而造成其因果关系比例显著高于 L1。另外,L2 的时态一致性比 L1 要低,说明 L2 驾驭时态的能力与 L1 相比仍有较大差距。

Table 9. Word frequency (Per Thousand Word) and ratio statistics of situation model cohesion
表 9. 情景模型衔接词频(每千词)和比例统计

情景模型衔接 (平均值)	因果动词 (词频)	因果动词及因果虚词 (词频)	意向动词 (词频)	因果关系 (比例)	意向性 (比例)	时态一致性 (比例)
L2 引言语料库	20.51	33.3	12.78	0.60	1.29	0.75
L1 引言对比语 料库	21.25	29.72	13.62	0.38	1.35	0.83
p 值	0.474	0.015	0.394	0.001	0.655	0.000

因果关系对于构建事件和行动之间的因果关系很重要。研究发现,L2 比 L1 运用了更多表示因果关系的虚词。例 13~14 表明,比起仅使用表示因果关系的动词来表明因果关系,如: cause, bring about, result in, result from 等,L2 倾向于将因果动词(make)与因果虚词(thus, in order to)结合在一起使用。因果虚词可以帮助读者更好地理解因果关系,但同时也造成因果关系比例过高,在隐性衔接的使用上逊色于 L1。Crossley 和 McNamara 提出,因果关系比例较高的文本在建立连贯的文本表征时使用更多的因果虚词,这样的文本衔接比较随意;而如果因果动词较多,在缺乏因果虚词辅助的情况下,读者反而需要主动思考并深度理解文本中的因果关系。这也能够解释为什么理解 L1 的引言要更吃力,而 L2 的引言相对更好理解。

例 13 (L2): Thus, the fuel reaction rate mainly depends on the mixing rate of fuel and air, and easily formed local high-temperature and fuel-rich regions make it difficult to avoid high NOx and soot emissions.

例 14 (L2): In order to make it possible that more energy is used to melt the wire while only a small fraction of energy is used to melt the parent material, the DE-GMAW is proposed.

学习者出现上述问题可能与写作教学密切相关。从社会建构主义的角度来看,引言中的文献引用是科研话语体系的一部分,具有推介功能,是说服读者、增加论文可信度和权威性的工具[21]。作者一般要在回顾既往研究的基础上找到研究空缺,进而提出本篇论文的研究价值,从而导致了更高的因果关联性。

时态一致性作为隐性衔接,是读者理解引言的关键手段,而 L2 引言中出现了多处时态一致性问题。其中,学生在主从复合句和相邻句子时态一致方面出现较多问题。如例 15 应把宾语从句的动词分别改为 had 和 was,例 16 也应该把宾语从句中的 has 改为 had,从而使主从复合句时态一致。而例 15 还应该把第二个句子的谓语动词改成 were,例 17 应该把 was 改为 is,使相邻句子时态一致。

例 15 (L2): The experiment showed that the catalyst has excellent electrocatalytic carbon dioxide reduction selectivity, and it is in long-term electrolysis. After that, the original morphology and ruthenium distribution are still maintained, indicating good stability.

例 16 (L2): Dastidar and Amyotte studied the inhibition of monoammonium phosphate and sodium bicarbonate for aluminum dust explosion, and the results showed that the former has better inhibition effect compared with the latter one.

例 17 (L2): In order to make it possible that more energy is used to melt the wire while only a small fraction of energy is used to melt the parent material, the DE-GMAW is proposed. A portion of current was decoupled by employing a bypass filler wire.

运用好时态与学生的英语水平密切相关,可以理解 L2 运用时态的能力相对薄弱,时态一致性远不如 L1。此外,引言篇幅较长,又要实现推介功能,包括确立研究领域、设置合适的研究地位和占领研究地位等不同语步[22],其使用到的时态种类十分丰富,对 L2 也造成一定困难。

5. 结论

本文拓展了衔接理论框架,基于自建语料库,结合 Coh-Metrix 指标,从显性衔接和隐性衔接两个方面分析了 L2 的引言衔接手段。研究表明,L2 运用衔接的能力与 L1 有所差别,存在衔接手段使用不足或使用过度等问题。1) 就显性衔接而言,语法衔接方面,L2 连词使用过度,使用所指总数及第一人称复数不足,而频繁使用被动语态或采取客观事物做主语的方式以提高论文的客观性;词汇衔接方面,无论是局部重复和整体重复,L2 均存在过度重复的问题,其结果可以使引言更容易理解,但却会导致文章用词单一、语句结构单调等问题。2) 就隐性衔接而言,语义衔接方面,L2 句间语义相似度过高,旧信息占比过高,存在语义过度重复的问题;情景模型衔接方面,过度使用因果虚词,而时态一致性不足。以上特征反映出 L2 在隐性衔接的方面存在一定困难,缺乏一定的写作技巧,因此,作为学术新手,L2 有必要了解引言的衔接手段特征,尤其是隐性衔接手段,掌握引言对话与交流的写作规范。

研究表明,影响学习者引言部分运用衔接手段的因素有很多,如:英语水平、写作教学、写作意图以及汉语负迁移等。因此,在学术写作教学中,教师应关注衔接问题,引导学生在学术引言写作中实现更好的连贯和衔接。限于篇幅,本研究选取的语料有限,且对引言衔接手段的研究仅局限于理工科类的化学和能源类两大学科,未来的研究将继续扩大语料规模,拓展研究学科,逐步推进中国学习者学术论文写作衔接手段的研究。

基金项目

大连理工大学研究生教改基金:“博士生英语学术论文写作的衔接手段研究”的资助。

参考文献

- [1] Halliday, M.A.K. and Hasan, R. (1976) *Cohesion in English*. Longman, New York.
- [2] 胡壮麟. 有关语篇衔接理论多层次模式的思考[J]. 外国语(上海外国语大学学报), 1996(1): 1-8.
- [3] 张德禄, 刘汝山. 语篇连贯与衔接理论的发展及应用[M]. 上海: 上海外语英语教学出版社, 2018.
- [4] Crossley, S.A. and McNamara, D.S. (2009) Computational Assessment of Lexical Differences in L1 and L2 Writing. *Journal of Second Language Writing*, **18**, 119-135. <https://doi.org/10.1016/j.jslw.2009.02.002>
- [5] 王晶晶, 姜峰. 中国理工科博士生学术论文写作立场建构研究[J]. 外语界, 2019(3): 23-31.
- [6] 王晶晶, 吕中舌. 理工科博士生学术英语写作中的作者自我指称语研究[J]. 外语界, 2017(2): 89-96.
- [7] 刘应亮, 陈洋. 中美学生硕士论文写作中立场标记语对比研究[J]. 中国外语, 2020, 17(2): 81-89.

-
- [8] McNamara, D.S., Graesser, A.C., McCarthy, P.M. and Cai, Z.G. (2014) Automated Evaluation of Text and Discourse with Coh-Metrix. Cambridge University Press, Cambridge. <https://doi.org/10.1017/CBO9780511894664>
- [9] 唐咏雪, 夏伟蓉. 科技论文英文摘要中逻辑衔接的运用分析[J]. 云南民族大学学报(哲学社会科学版), 2009, 26(6): 154-157.
- [10] 刘辉. 学术期刊方法部分的指称性衔接分析[J]. 外语学刊, 2018(5): 67-72.
- [11] 何中清, 闫煜菲. 跨学科学术论文引言中的元话语对比研究[J]. 外语教学, 2021, 42(5): 26-30.
- [12] Swales, J.M. and Feak, C.B. (2012) Academic Writing for Graduate Students: Essential Tasks and Skills. Michigan University Press, Ann Arbor. <https://doi.org/10.3998/mpub.2173936>
- [13] Reid, J.R. (1992) A Computer Text Analysis of Four Cohesion Devices in English Discourse by Native and Nonnative Writers. *Journal of Second Language Writing*, 1, 79-107. [https://doi.org/10.1016/1060-3743\(92\)90010-M](https://doi.org/10.1016/1060-3743(92)90010-M)
- [14] Cain, K. and Nash, H.M. (2011) The Influence of Connectives on Young Readers' Processing and Comprehension of Text. *Journal of Educational Psychology*, 103, 429-441. <https://doi.org/10.1037/a0022824>
- [15] 王月丽, 徐宏亮. 中国英语学习者学术写作中第一人称使用发展特征与身份构建研究[J]. 外语教育研究前沿, 2019, 2(3): 58-64.
- [16] Harwood, N. (2005) "Nowhere Has Anyone Attempted... In This Article I Aim to Do Just That": A Corpus-Based Study of Self-Promotional I and We in Academic Writing across Four Disciplines. *Journal of Pragmatics*, 37, 1207-1231. <https://doi.org/10.1016/j.pragma.2005.01.012>
- [17] Wales, K. (1996) Personal Pronouns in Present-Day English. Cambridge University Press, Cambridge.
- [18] 许双如. 语篇衔接与英语习得程度调查分析[J]. 广州大学学报(社会科学版), 2004, 3(5): 41-44.
- [19] 王力. 中国语法理论[M]. 济南: 山东英语教学出版社, 1984.
- [20] 连淑能. 英汉对比研究[M]. 北京: 高等英语教学出版社, 1993.
- [21] 刘宇, 李武. 引文评价合法性研究——基于引文功能和引用动机研究的综合考察[J]. 南京大学学报(哲学·人文科学·社会科学版), 2013, 50(6): 137-148+157.
- [22] Swales, J.M. (1990) Genre Analysis: English in Academic and Research Settings. Cambridge University Press, Cambridge.