

北理工外语学术文库

中国德语专业大学生作文中的 篇章衔接与连贯：

一项基于语料库与潜伏语义分析的研究

Cohesion and Coherence in the Essays of
German Major Students in China: A Corpus- and LSA-based Study

刘磊 著 ■

外语教学与研究出版社
FOREIGN LANGUAGE TEACHING AND RESEARCH PRESS

中国德语专业大学生作文中的 篇章衔接与连贯：

一项基于语料库与潜伏语义分析的研究

Cohesion and Coherence in the Essays of
German Major Students in China: A Corpus- and LSA-based Study

刘磊 著 ■

图书在版编目 (CIP) 数据

中国德语专业大学生作文中的篇章衔接与连贯：一项基于语料库与潜伏语义分析的研究 / 刘磊著. — 北京：外语教学与研究出版社，2016.6
(北理工外语学术文库)
ISBN 978-7-5135-7771-7

I. ①中… II. ①刘… III. ①德语—写作—研究 IV. ①H335

中国版本图书馆 CIP 数据核字 (2016) 第 152174 号

出 版 人 蔡剑峰
项目负责 孔乃卓
责任编辑 李婉婧
封面设计 郭 子
版式设计 付玉梅
出版发行 外语教学与研究出版社
社 址 北京市西三环北路 19 号 (100089)
网 址 <http://www.fltrp.com>
印 刷 北京九州迅驰传媒文化有限公司
开 本 650×980 1/16
印 张 19.75
版 次 2016 年 7 月第 1 版 2016 年 7 月第 1 次印刷
书 号 ISBN 978-7-5135-7771-7
定 价 54.90 元

购书咨询：(010) 88819926 电子邮箱：club@fltrp.com
外研书店：<https://waiyants.tmall.com>
凡印刷、装订质量问题，请联系我社印制部
联系电话：(010) 61207896 电子邮箱：zhijian@fltrp.com
凡侵权、盗版书籍线索，请联系我社法律事务部
举报电话：(010) 88817519 电子邮箱：banquan@fltrp.com
法律顾问：立方律师事务所 刘旭东律师
中咨律师事务所 殷 斌律师
物料号：277710001

北理工外语学术文库

主编：李京廉 副主编：姜爱红

编委会

主任：王克非 北京外国语大学

委员（按拼音顺序）：

曹大峰 北京日本学研究中心

陈大明 北京理工大学

程朝翔 北京大学

姜爱红 北京理工大学

李京廉 北京理工大学

刘文杰 北京理工大学

罗选民 清华大学

苗兴伟 北京师范大学

钱敏汝 北京外国语大学

王玉雯 北京理工大学

叶云屏 北京理工大学

郑书九 北京外国语大学

总 序

百年大计，教育为本；国家兴旺，人才为基。如何担负起培养国家栋梁之才的责任，是大学和高校教师必须认真思考和回答的问题。十九世纪下半叶，德国著名学者、教育改革家威廉·冯·洪堡（Wilhelm von Humboldt）提出现代大学的两个重要使命：教学和科研。虽然二十世纪以后大学又增添了社会服务功能，但教学与科研是大学服务社会的基础和手段。教学是大学满足人们求学深造的愿望以及社会对各种专业人才需求的最主要、最有效的方式。而要将既有知识和学术前沿不断地传授给学生，就需要教师紧跟学科发展趋势，不断更新教学内容，改进教学方法。教师也只有从事科学研究才能保证学术水平不断提高。如果没有一定的学术水平，没有坚实的理论基础和足够的科研积累，则难以胜任大学的教学工作。诚如雅思贝尔斯所言：“最好的研究者才是最优良的教师。只有这样的研究者才能带领人们接触真正的求知过程，乃至科学的精神。”（雅思贝尔斯·《什么是教育》·北京：生活·读书·新知三联书店，1991：152）

正是秉承上述大学精神，我院历来重视学科专业建设，鼓励教师积极从事教学科研工作，努力取得高水平研究成果。《北理工外语学术文库》就是学院教师多年教学实践和科学研究积累结出的硕果，是适应高等教育发展形势，不断提高自身素质，提升教学质量与科研水平的集中体现。

《北理工外语学术文库》包含二十余部学术专著，涉及理论语言学、

应用语言学、语言与文化以及文学等外国语言文学研究的主干领域，涵盖英语、德语、日语等不同语种，既有理论探索、又有实证研究，既有语言构造、文本、语篇的微观分析，又有思想、文化等宏观研究视角。文库作者多为学院青年教师，是学院教学科研的中坚力量和未来发展的希望所在。我们相信，《北理工外语学术文库》的出版，将提升学院教师的科研积极性，推动学院教学科研再上新台阶。

学术文库的出版得到北京理工大学学科建设经费的资助，得到学校领导和相关职能部门的大力支持，我们对此表示衷心感谢。同时，我们也十分感谢文库编委会在出版谋划、选题和内容审校过程中所付出的辛勤劳动，感谢外语教学与研究出版社同仁的鼎力相助。在文库出版之际，我们真切地期望得到广大读者、同行和专家的关注和指正！

北京理工大学外国语学院

二零一四年十月二十八日

序

刘磊和我，亦师亦友，也是同行。所以当他约我为他的这本专著写序的时候，我毫不犹豫，欣然应允。

国内目前有近 110 所高校设有德语语言文学专业，每年有 5000 多人参加德语四级考试，近 3000 人参加德语八级考试。这本书是一项以中国高校德语专业大学生的作文作为研究对象、基于语料库的研究成果，是刘磊的第一本专著，也是国内该领域第一部正式出版的研究成果。因此，它对于个人以及中国德语界的意义不言而喻。

这本书里既有刘磊求学期间的学术积累，也有他从教后的观察、思考和探索，从中我们能够清晰地感到刘磊是一个敏于思考、孜孜不倦、内心坚定的人；其中尤其让人难忘的是他对语言学这门学科的挚爱以及不懈的追求。

作为国内德语语言文学专业的一名后起之秀，刘磊的科研之路才刚刚开始，在此我衷心祝愿他在这条道路上越走越远，享受科研的乐趣，取得卓越的成绩。

是为序。

缪雨露

北京外国语大学德语系

前 言

计算机语料库所提供的大量真实语料与相关软件相结合,使语言学研究产生了丰硕的成果,也对外语教学起到了积极的推动作用。随着国内首个德语学习者语料库的建立,基于它的语言学研究也逐渐出现。基于计算机技术,对中国德语学习者语料库中的作文进行研究,能够为中国德语写作教学的完善起到重要的作用。因此,本书选取中国德语学习者语料库中PGG子库和自建德国母语者语料库中的作文作为语料,借助计算机技术对两者作文中的衔接与连贯进行研究。研究目的在于:一方面,使基于计算机技术的篇章语言学研究范畴不断拓展;另一方面,为中国德语专业教师和学习者提供建议。

本书分为六个部分,即:前言、理论基础、语料加工处理原则与技术、实证分析、教学建议、结束语。前言部分(第一部分)主要介绍了本书的选题缘由、研究对象的界定、语料的选择、国内外研究现状、理论基础和研究方法以及结构。本书的理论部分(第二部分)首先探讨了不同视角的篇章观并将钱敏汝的篇章语用学视角下的动态篇章观作为基本出发点。然后,在相关学者衔接与连贯理念的基础上结合中国德语学习者作文的语言特点,确定了本研究的衔接与连贯类别。其次,对语料库语言学哲学基础、研究方法、语料库定义进行了阐述。接下来,又论述了潜伏语义分析的基本原理、实现、应用、衔接和连贯与潜伏语义分析之间的关联。最后,从计算机辅助语言教学、篇章与人和世界的关联以及建构主义理论的角度出发,为计算机辅助德语写作教学构建了理论基础。在本书的第三部分,阐述了中国德语学习者和德国母语者语料数

据加工处理原则与技术。在第四部分即实证分析部分,对中国德语学习者和德国母语者作文中的衔接手段的使用、局部连贯和整体连贯进行了分析。在本书的第五部分,主要为中国德语专业教师以及学习者提出了计算机辅助德语写作教学的建议。第六部分为结束语。

在实证研究的基础上,本书得出了如下结论:

(1) 在衔接手段使用的总体分布上,中国德语学习者和德国母语者是相同的,即:使用重复最为频繁,其次是关联词和替代,使用替换最少。中国德语学习者作文中完全重复、同义替换、代词替代、并列连词(对立关系)、从属连词(目的关系、dass/ob)的使用显著多于德国母语者;部分重复、代副词替代、并列连词(叠加、选言关系)、从属连词(结果、让步关系)、连接副词(并列、时序、让步关系)的使用显著少于德国母语者;而上下文替换、并列连词(阐释、因果关系)、从属连词(因果、条件、情态、时序关系)、连接副词(因果、对立关系)的使用与德国母语者没有明显差异。通过相关性检验发现,在所研究的24种衔接手段中,仅有4种(同义替换、代副词、并列连词阐释、连接副词并列)能够预测作文质量,而其他20种并不能预测作文质量。

(2) 中国德语学习者和德国母语者作文相邻句子间、相邻段落间、全部句子间的LSA余弦均值不存在显著性差异,而中国德语学习者作文全部段落间LSA余弦均值要显著低于德国母语者。这也就说明,中国德语学习者和德国母语者作文在句子间的局部连贯性、整体连贯性和段落间的局部连贯性上不存在显著性差异,而中国德语学习者作文段落间的整体连贯性要显著低于德国母语者。通过相关性检验发现,中国德语学习者作文中相邻段落间、全部段落间的LSA余弦均值能够预测作文质量。

(3) 在实证研究的基础上,笔者为中国德语专业教师和学习者提供了计算机辅助德语写作教学的建议。一方面,中国德语专业教师可以借助计算机软件来自主创建微型文本、探索中国德语学习者作文的语言特点、研发德语测试试题、探索学生作文的自动评判。另一方面,中国德语学习者可以借助计算机软件自主建构语言层面的知识,自主建构文化层面的知识,使用第四代语料库工具进行自主学习。

综上所述,本书的学术贡献以及创新尝试主要体现在以下几个方面:

(1) 本书将篇章语用学视角下的动态篇章观以及篇章和人与世界的关联和语料库语言学、潜语义义分析相结合,这是对篇章语用学的拓展和发展。借助计算机技术对自然语言进行研究,也不断由词汇、句子层

面拓展到篇章层面，这展示了未来语言学研究的重要发展方向。

(2) 本书首次尝试使用潜伏语义分析对德语学习者作文中的连贯进行研究。以往对篇章连贯的研究基本都停留在篇章表层结构之中，本书的尝试对篇章连贯测量标准的确立具有一定的贡献。

(3) 本书结合计算机辅助语言教学、篇章和人与世界的关联、建构主义构建了计算机辅助德语写作教学的理论框架。篇章语用学的相关理念除了应用于语言研究中，同样可以应用到德语写作教学之中。

(4) 由于目前在德国还没有母语者描述型篇章语料库，因此笔者建立了首个小型德国母语者描述型篇章语料库。这样的德国母语者语料库不仅能够为基于语料库的研究做参照，而且能够作为语言材料直接应用到中国德语写作教学之中。

(5) 本书所提供的计算机辅助德语写作教学的建议能够提高德语教师的工作效率，并且改善德语学习者自主学习的能力。

目 录

图表目录	xii
------------	-----

第1章 引言	1
--------------	---

1.1 选题缘由	1
1.2 研究对象的界定	3
1.3 语料的选择	5
1.4 国内外研究纵览	10
1.4.1 篇章衔接与连贯的研究现状	10
1.4.2 基于语料库的研究现状	12
1.4.3 基于潜伏语义分析的研究现状	14
1.5 理论基础和研究方法	15
1.6 本书结构	17

第2章 基本概念与理论基础	19
---------------------	----

2.1 篇章语言学	19
2.1.1 篇章定义	19
2.1.2 篇章衔接与衔接手段	26
2.1.3 篇章连贯	31
2.1.4 衔接与连贯的关系	38

2.2	语料库语言学和语料库	41
2.2.1	语料库语言学	41
2.2.2	语料库	50
2.3	潜伏语义分析理论	57
2.3.1	潜伏语义分析的基本原理	57
2.3.2	潜伏语义分析的实现	62
2.3.3	潜伏语义分析的应用	72
2.3.4	篇章衔接和连贯与潜伏语义分析	74
2.4	计算机辅助德语写作教学	79
2.4.1	计算机辅助语言教学	80
2.4.2	篇章语用学视角下篇章与人和世界的关联	82
2.4.3	建构主义教学理论	84
2.4.4	计算机辅助教学、篇章语用学、建构主义与 德语写作教学的关联	86
2.5	本章小结	87
第3章	中国德语学习者和德国母语者语料加工处理原则与 技术	89
3.1	语料的标注	89
3.2	语料的切分	96
3.2.1	句子的切分	96
3.2.2	段落的切分	100
3.3	数据分析软件及数据统计	102
3.3.1	软件PowerGREP以及正则表达式编写	102
3.3.2	软件Text to SVD	104
3.3.3	数据统计	104
3.4	本章小结	105
第4章	中国德语学习者和德国母语者语料之实证分析	107
4.1	衔接	107
4.1.1	中国德语学习者和德国母语者作文中衔接手段 使用差异性研究	107
4.1.2	中国德语学习者和德国母语者作文中各种衔接手段 使用评析	125

4.1.3	中国德语学习者作文中衔接手段使用与 作文质量之间相关性检验	168
4.2	连贯	172
4.2.1	中国德语学习者和德国母语者作文中连贯性差异性 研究	173
4.2.2	中国德语学习者和德国母语者作文连贯性评析	178
4.2.3	中国德语学习者作文连贯性与作文质量相关性检验 ...	237
4.3	本章小结	238
第5章	计算机辅助德语写作教学之建议	241
5.1	对实证分析结果的思考	241
5.2	本研究对德语写作教学的启示	243
5.2.1	德语专业教师层面	243
5.2.2	德语学习者层面	255
5.3	本章小结	266
第6章	结束语	267
6.1	研究结论	267
6.2	创新尝试	275
6.3	局限性和对未来研究的展望	277
附录	德国母语者语料库建设——写作指南	279
参考文献		282

图表目录

表 1-1	中国德语学习者语料库一览表.....	5
表 1-2	中国德语学习者语料库 2009 年 PGG 子库作文题目及引言.....	7
表 1-3	中国德语学习者语料抽样以及分类数据明细.....	8
表 1-4	德语母语学习者语料数据明细.....	10
图 2-1	人 $\leftarrow$ 篇章 $\rightarrow$ 世界的三元关系 1.....	24
图 2-2	人 $\leftarrow$ 篇章 $\rightarrow$ 世界的三元关系 2.....	26
表 2-1	原始矩阵样例.....	63
图 2-3	原始矩阵分解与重建.....	66
表 2-2	原始语料.....	67
表 2-3	原始矩阵 $\{X\}$	67
表 2-4	词语矩阵 $\{W\}$	68
表 2-5	对角矩阵 $\{S\}$	68
表 2-6	转置的文本矩阵 $\{P^t\}$	69
表 2-7	重建的矩阵 $\{X\}=\{W\} \times \{S\} \times \{P^t\}$	69
图 2-4	三维潜伏语义空间.....	70
图 2-5	相邻句子间余弦值.....	77
图 2-6	相邻段落间余弦值.....	78
图 2-7	全部句子间余弦值.....	78
图 2-8	全部段落间余弦值.....	78
图 2-9	篇章 $\leftarrow$ 人 $\rightarrow$ 世界之间的关系.....	82
图 2-10	人 $\leftarrow$ 篇章 $\rightarrow$ 世界之间的关系.....	82
图 2-11	篇章 $\leftarrow$ 世界 $\rightarrow$ 人之间的关系.....	83
表 3-1	各种衔接手段赋码集.....	90
图 3-1	AnnoTool 软件使用	92
图 3-2	原始文本 (09025ZW003)	99
图 3-3	在命令提示符中输入上述命令.....	99
图 3-4	文本中各个句子单独成为段落.....	100

图 3-5	输入命令及切分后的文本.....	100
图 3-6	原始文本 (HU12).....	101
图 3-7	输入命令及切分后的段落.....	101
表 3-2	检索各种衔接手段的正则表达式.....	102
图 3-8	Text to SVD 操作界面.....	104
表 4-1	中国德语学习者和德国母语者衔接手段使用总体分布.....	108
表 4-2	中国德语学习者与德国母语者衔接手段使用总体分布 (标准化频率).....	108
图 4-1	中国德语学习者与德国母语者衔接手段使用总体分布 对比图.....	109
表 4-3	中国德语学习者和德国母语者作文中完全重复百分比 独立样本 t 检验.....	110
表 4-4	中国德语学习者和德国母语者作文中部分重复百分比 独立样本 t 检验.....	110
表 4-5	中国德语学习者和德国母语者作文中同义替换百分比 独立样本 t 检验.....	112
表 4-6	中国德语学习者和德国母语者作文中上下义替换百分比 独立样本 t 检验.....	112
表 4-7	中国德语学习者和德国母语者作文中代词百分比 独立样本 t 检验.....	113
表 4-8	中国德语学习者和德国母语者作文中代副词百分比 独立样本 t 检验.....	113
表 4-9	中国德语学习者和德国母语者作文中并列连词 (叠加) 百分比独立样本 t 检验.....	115
表 4-10	中国德语学习者和德国母语者作文中并列连词 (对立) 百分比独立样本 t 检验.....	116
表 4-11	中国德语学习者和德国母语者作文中并列连词 (选言) 百分比独立样本 t 检验.....	116
表 4-12	中国德语学习者和德国母语者作文中并列连词 (阐释) 百分比独立样本 t 检验.....	116
表 4-13	中国德语学习者和德国母语者作文中并列连词 (因果) 百分比独立样本 t 检验.....	116
表 4-14	中国德语学习者和德国母语者作文中从属连词 (时序) 百分比独立样本 t 检验.....	118

表 4-15	中国德语学习者和德国母语者作文中从属连词（因果） 百分比独立样本 t 检验	118
表 4-16	中国德语学习者和德国母语者作文中从属连词（结果） 百分比独立样本 t 检验	119
表 4-17	中国德语学习者和德国母语者作文中从属连词（条件） 百分比独立样本 t 检验	119
表 4-18	中国德语学习者和德国母语者作文中从属连词（情态） 百分比独立样本 t 检验	119
表 4-19	中国德语学习者和德国母语者作文中从属连词（让步） 百分比独立样本 t 检验	120
表 4-20	中国德语学习者和德国母语者作文中从属连词（目的） 百分比独立样本 t 检验	120
表 4-21	中国德语学习者和德国母语者作文中从属连词（dass/ob） 百分比独立样本 t 检验	120
表 4-22	中国德语学习者和德国母语者作文中连接副词（并列） 百分比独立样本 t 检验	123
表 4-23	中国德语学习者和德国母语者作文中连接副词（时序） 百分比独立样本 t 检验	123
表 4-24	中国德语学习者和德国母语者作文中连接副词（因果） 百分比独立样本 t 检验	123
表 4-25	中国德语学习者和德国母语者作文中连接副词（让步） 百分比独立样本 t 检验	123
表 4-26	中国德语学习者和德国母语者作文中连接副词（对立） 百分比独立样本 t 检验	124
表 4-27	中国德语学习者和德国母语者实现完全重复的语言手段 分布	126
图 4-2	中国德语学习者和德国母语者实现完全重复的语言手段 分布对比	128
表 4-28	中国德语学习者和德国母语者实现部分重复的语言手段 分布	131
图 4-3	中国德语学习者和德国母语者实现部分重复的语言手段 分布对比	134
表 4-29	中国德语学习者和德国母语者实现替换的语言手段分布	136
图 4-4	中国德语学习者和德国母语者实现替换的语言手段分布	

	对比.....	138
表 4-30	中国德语学习者和德国母语者实现替代的语言手段分布.....	140
图 4-5	中国德语学习者和德国母语者作文中各种代词、代副词 替代分布对比.....	141
表 4-31	代词和代副词对数似然比检验结果.....	144
表 4-32	中国德语学习者和德国母语者各类关联词使用分布.....	148
图 4-6	中国德语学习者和德国母语者作文中各种并列连词 分布对比.....	148
表 4-33	并列连词对数似然比检验结果.....	152
图 4-7	中国德语学习者和德国母语者作文中各种从属连词分布 对比.....	154
表 4-34	从属连词对数似然比检验结果.....	159
图 4-8	中国德语学习者和德国母语者作文中各种连接副词分布 对比.....	162
表 4-35	连接副词对数似然比检验结果.....	166
表 4-36	各种衔接手段与作文质量之间相关性检验.....	168
表 4-37	中国德语学习者和德国母语者作文相邻句子间 LSA 余弦 均值独立样本 t 检验	173
表 4-38	中国德语学习者和德国母语者作文相邻段落间 LSA 余弦 均值独立样本 t 检验	174
表 4-39	中国德语学习者和德国母语者作文全部句子间 LSA 余弦 均值独立样本 t 检验	176
表 4-40	中国德语学习者和德国母语者作文全部段落间 LSA 余弦 均值独立样本 t 检验	176
表 4-41	中国德语学习者作文 09022ZW046 相邻句子间余弦值	179
表 4-42	中国德语学习者作文 09071ZW054 相邻句子间余弦值	181
表 4-43	中国德语学习者作文 09018ZW229 相邻句子间余弦值	184
表 4-44	德国母语者作文 HU37 相邻句子间余弦值.....	186
表 4-45	德国母语者作文 HU44 相邻句子间余弦值.....	188
表 4-46	德国母语者作文 HU52 相邻句子间余弦值.....	191
表 4-47	中国德语学习者作文 09020ZW042 相邻段落间余弦值	193
表 4-48	中国德语学习者作文 09023ZW017 相邻段落间余弦值	195
表 4-49	中国德语学习者作文 09076ZW003 相邻段落间余弦值	196
表 4-50	德国母语者作文 HU18 相邻段落间余弦值.....	198