

# 人文社科学术论文语言变异的多维度分析\*

袁亮杰<sup>1</sup>, 王治敏<sup>1</sup>, 朱宇<sup>2†</sup>

<sup>1</sup>北京语言大学汉语国际教育研究院, 北京100083

<sup>2</sup>厦门大学国际中文教育学院, 福建厦门361102

{yljlarry,wangzm000}@qq.com, zhuyu@xmu.edu.cn

## 摘要

通过自建人文和社科领域中文学术期刊论文语料库(逾920万字),运用多维度分析法对111项语言特征的频次数据进行因子分析和维度识别,发现人文社科领域学术文具有7个维度的语言特征共现模式:描述性vs.阐释性、概念判断vs.行为再现、铺陈与发展、已然性表述、计数与测量、模糊性表达、顺序与连接。进而,对语料在上述各维度的量化表现施以统计检验和聚类分析,发现学术语体在人文与社科两大领域的语言变异显著体现于除“计数与测量”、“顺序与连接”以外的其他5个维度;人文领域和社科领域内部学科的语言变异,各在6个维度上存在显著差异。本研究为学术汉语写作、汉语语体语法等提供一定启示。

**关键词:** 语言变异; 多维度分析; 学术汉语

## A multi-dimensional analysis of register variations in Chinese academic papers of Humanities and Social Sciences

Yuan Liangjie<sup>1</sup>, Wang Zhimin<sup>1</sup>, Zhu Yu<sup>2</sup>

<sup>1</sup>Research Institute of International Chinese Language Education,  
Beijing Language and Culture University, Beijing 100083

<sup>2</sup>Chinese International Education College,  
Xiamen University, Xiamen, Fujian 361102

{yljlarry,wangzm000}@qq.com, zhuyu@xmu.edu.cn

## Abstract

Through self-built corpus of Chinese academic journal papers in the field of humanities and social sciences (more than 9.2 million words), using Multi-Dimensional Analysis to conduct factor analysis and dimension identification on the frequency data of 111 language features, it is found that Chinese academic papers in the field of humanities and social sciences have 7 co-occurrence patterns: descriptive vs. interpretative, conceptual judgment vs. behavior reappearance, spreading and development, existentiality statement, counting and measuring, expression of hedging, sequence and connection. Furthermore, one-way MANOVA and cluster analysis reveals variations of academic Chinese journals between humanities and social sciences exist in 5 dimensions, also reveals variations exist within 6 dimensions among the disciplines of humanities and social science. This research provides more objective and pertinent suggestions for academic Chinese writing, Stylistic-Register Grammar, etc.

**Keywords:** register variation, Multi-Dimensional Analysis, academic Chinese

\* 基金项目: 国家社科基金重大项目(18ZDA295); 国家社科基金后期资助项目(21FYYB010); 教育部人文社科研究规划基金项目(21YJA740058); 陕西省教育厅人文社科专项项目(20JK0086); 陕西理工大学校级科研项目(SLGKY2024)

† 通讯作者 corresponding author

## 1 引言

学术期刊论文是学术成果的重要载体，其语体功能是准确记述自然、社会及人类思维现象，严密论证其内在规律等(王德春, 2004)，与文学、新闻等相区分。语体动因的功能类型塑造了语法(方梅, 2007)，而恰切的语体观察，是说明语法规律的最佳途径(张伯江, 2012)。语域是与特定使用情境相联系的一种语言变体(Biber and Conrad, 2019)。在学术语体内部，不同学科领域的语言使用情境不同，从而形成各具差异的学科语域。考察不同学科语域的语言特征及其变异情况，能够概括汉语学术语体的语法规律。

随着多维度分析法(Multi-Dimensional Analysis)的日臻完善，近年来学者不断借鉴并跟随其首倡者Douglas Biber(1985; 1986; 1988; 1995; 2006; 2014)的研究路径，借助大规模真实语料库和因子分析的统计方法，得到语言特征在语料中的若干聚合模式(即维度)，从而以量化视角探索跨语体/语域的语言变异情况(Biber and Kurjian, 2006; Biber et al., 2016; Biber and Egbert, 2016; Biber and Conrad, 2019)。雷秀云和杨惠中(2001)以及武姜生(2001)是国内较早系统介绍多维度分析法的文献，并籍此引领中国学者对学术英语和其他书面英语的语言变异情况进行描写(武姜生, 2004; 潘, 2012; 胡春雨, 谭金琳, 2020)。

在现代汉语语言变异的多维度分析方面，Zhang(2012)对汉语书面语的语言变异情况进行考察；朱晓楠(2014)考察16个汉语口语语域并提取出现代汉语语言变异的5个维度；刘艳春(2019)识别出汉语语体的7个变异维度；范楚琳和刘颖(2020)采用随机森林和k-means聚类算法定量比较了鲁迅三种文体的语言特征；朱宇和胡晓丹(2021)定量考察141项连词的共现模式，创新汉语虚词的量化研究路径。这些文献将多维度分析法运用于汉语研究和自然语言处理领域，是对汉语语体和语言变异进行实证考察的有益尝试，拓展了汉语语言变异研究的视野。但总体来看，国内学界罕见运用多维度分析法对学术汉语进行中观和微观视角下的专门描写与研究。因此，本文基于大规模中文学术期刊语料，尝试分领域、分学科考察学术汉语书面语的语言变异情况。研究问题如下：

- (1) 在人文领域和社科领域之间，中文学术期刊论文的语言变异有哪些异同？
- (2) 在人文领域内部和社科领域内部，中文学术期刊论文的语言变异各有哪些异同？

## 2 研究设计

### 2.1 语料库建设

为体现样本的典型性和代表性，本文选取CSSCI各学科排名前两位的期刊，建立中文学术期刊语料库(人文社科子库)。其中，人文领域选取中国文学、历史学和哲学，社科领域选取经济学、社会学和法学，共六门学科。具体抽样方法为：确定CSSCI上述六门学科排名前两位的12种期刊，记录其在相同时段内发表的全部学术论文(不含简讯、书讯、征稿启事等)，从各学科随机抽取25%作为研究语料，共计970篇，约920万字，见表1。

### 2.2 语言特征体系建立

本文结合现代汉语特点，参考语体研究成果，建立现代汉语语言特征体系，共21类111项(详见附录A)，涉及现代汉语词汇、语法、韵律等多个层面。该特征体系内的语言特征并非独立，而是存在交叉重叠、容纳包含关系。例如名词中既有按使用频次的分类，又有按抽象或具体特征的分类，还有按常用领域的分类等。旨在从多元视角，多方位审视学术汉语语言变异情况。

### 2.3 语料清洗和处理

为确保数据准确性，删除各篇学术论文的作者简介、基金项目、图表、附录、参考文献、英文摘要、英文关键词等内容。使用NLPIR汉语分词系统<sup>0</sup>对语料进行分词和词性标注。使用AntConc软件和自编程序Text Analysis对每篇文本中语言特征的频数进行检索与统计，并做归一化处理，得到每项语言特征在每篇文本内每千字的出现频率，作为语言特征频率数据。

©2022 中国计算语言学大会

根据《Creative Commons Attribution 4.0 International License》许可出版

<sup>0</sup><https://github.com/NLPIR-team/NLPIR>

领域	学科	选取期刊	篇数	字数
人文	中国文学	《文学评论》	104	1 327 487
		《文学遗产》	68	811 081
	历史学	《历史研究》	57	924 922
		《近代史研究》	41	721 352
	哲学	《哲学研究》	160	507 775
《哲学动态》		130	508 306	
社科	经济学	《经济研究》	109	684 100
		《世界经济》	68	755 321
	社会学	《社会学研究》	47	713 853
		《中国人口科学》	53	440 443
	法学	《中国法学》	72	922 832
		《法学研究》	61	874 337
总计			970	9 200 000

表 1: 语料库构成及规模

## 2.4 因子分析与维度建立

使用SPSS 22.0软件对语言特征频率数据进行因子分析，建立语言特征维度。KMO取样适切性量数0.884，Bartlett球形检验显著性 $p < 0.001$  (Approx.  $\chi^2 = 57813.725$ ,  $df = 4656$ )，表明数据适合进行因子分析。运用未加权的最小二乘法提取因子，并以Promax斜交方式进行因子旋转。经过对比研判和综合考量，参考碎石图曲线，选取前7个因子建立维度（总方差的累积解释为42.913%<sup>1</sup>），排除载荷绝对值小于0.3的特征项，共得到60项语言特征及其载荷值。

## 3 学科语域文本的维度分析

通过因子分析，本文成功识别出人文社科中文期刊论文的7个维度语言特征集及其载荷值。按照Biber(1988),载荷值越高说明其对该维度的贡献越大，相关性越强。语言特征根据其载荷值的正负分为正向特征和负向特征，代表其在一维度中的两个方向。每个维度既可包含方向相同的一组特征集（均为正向或负向特征），也可包含方向不同的两组特征集（既有正向也有负向特征），且该两组特征集在某些特定语域中往往呈现互补分布。

进而，根据语言特征频率数据的标准分，按照各维度不同载荷值进行定量加权，得到每篇文本在7个维度上的维度分数。该分数是标准化之后的值，不受文本篇幅等影响，故能够在同一基础上进行量化比较。以下参照各维度语言特征及学科语域维度分均值，结合具体文本，分别对人文社科语域的7个维度进行分析和命名。

### 3.1 维度一：描述性vs.阐释性

维度一包含33项语言特征，见表2。其中，正向特征的数量和种类较多，如“嵌偶单音词”常见于状语，满足庄严语体需求(黄梅, 冯胜利, 2009)；“副词：中度常用”<sup>2</sup>、“形容词：非常用”“副词：最常用”等以形容词和副词为主，多用于描写和修饰；“动词‘有’，表存现”和“存现动词”常用在描述话语；“特指疑问句”、“第一人称代词‘我’”指代确定的人或事物。以上载荷值较高的语言特征均具有一定的描写和叙述功能。

负向特征中，“词长”和“平均句长”是对文本词句的度量，体现词汇和句子的复杂度和丰富度。“名动词”“名词：最常用”“抽象名词”和“集体名词”等承载较高信息密度。“形式动词”在书面

<sup>1</sup>关于本文中该数值的大小是否合理的问题，笔者曾向多维度分析法的首倡者、美国北亚利桑那大学英语系Dr. Douglas Biber教授发邮件请教，得到肯定答复。在此，特向Dr. Douglas Biber教授致以诚挚感谢！

<sup>2</sup>根据国家语委《现代汉语语料库词语分词类频率表》，本文将名词、动词、形容词、副词这四类词语，按照出现频率排序，各分为三个水平：最常用（前30%）、中度常用（31%-65%）、非常用（后35%）。该12项语言特征仅从频率角度划分，与其他语言特征存在交叉重叠。

维度	语言特征	载荷值	语言特征	载荷值
维度一	嵌偶单音词	0.974	存现动词	0.326
	语气词	0.855	假设复句	0.324
	其他第一人称代词	0.726	动词: 非常用	0.323
	动词“有”, 表存现	0.697	名词: 非常用	0.320
	目的复句	0.694	特指疑问句	0.320
	副词: 中度常用	0.671	第一人称代词“我”	0.312
	指示代词	0.587	状态词	0.307
	形容词: 非常用	0.544	集体名词	-0.313
	动词: 中度常用	0.476	抽象名词	-0.318
	副词: 非常用	0.457	形式动词	-0.372
	副词: 最常用	0.445	名词: 最常用	-0.392
	交流动词	0.421	助词“的”	-0.457
	并列复句	0.418	平均句长	-0.573
	形容词: 中度常用	0.407	合偶双音词	-0.586
	时间副词	0.399	名动词	-0.659
	增强语	0.341	词长	-0.930
	介词短语	0.328		

表 2: 维度一的语言特征及其载荷值

语中“经常用于庄重、典雅的文体”(刁晏斌, 2004), “合偶双音词”则是体现书面正式语体功能的语法结构手段(王永娜, 2015)。这些特征体现出较强的信息聚集, 通常用于阐说和解释。以下通过具体文本分析:

语料 (1) 节选自历史学期刊论文, 其中用不同符号标注出维度一的部分正向特征(形容词、副词、指示代词、介词)。该语料包含许多形容词和副词, 描述性强; 亦含表示对象、范围、方式的介词, 以保证描述表达的精确和严密(袁晖, 李熙宗, 2005)。

(1) 这样惨切与残败之余的社会生活感受, 我们读到《聊斋》故事, 即知其所描写的~~环境气氛绝不是偶然的胡思乱想~~。……具见其以己作为作品中人物之作, 即以自己真切深挚的情感意绪以为书中人物的描写。即在人物创作与描绘中, 把自己的思想感情化了进去。

语料 (2) 节选自经济学期刊论文, 标注出部分负向特征(名动词、名词: 最常用、合偶双音词)。该语料共两个句子, 平均句长71, 平均词长1.95, 实词占57.53%, 用以解释“企业生产效率分化”的作用。可见社科论文对概念和术语的运用较多, 解析力求准确、严谨, 具有较高信息负荷, 体现出较强的阐释功能。

(2) 企业之间生产效率的分化, 进而激发了产业层面的重组和跨企业的资源再配置过程: 一方面, 生产率较低的企业逐渐收缩, 并最终退出市场; 另一方面, 新企业大规模进入, 生产率较高的企业也开始扩张。同时, 通过促进跨企业资源配置效率的改善, 银行部门的市场化也成为中国工业部门全要素生产率增长的重要源泉。

对六门学科和两大领域的文本维度分取均值, 绘制出图1, 可知: 人文领域, 三门学科在维度一的均值为正, 而社科领域三门学科均值为负, 呈现出明显分野。根据4.1节事后检验结果, 人文与社科两大领域及各领域内学科之间在维度一均具有显著性差异, 这说明学术期刊论文在维度一“描述性vs.阐释性”存在普遍的语言变异情况, 也是其最基本的功能。

### 3.2 维度二: 概念判断vs.行为再现

共有15项语言特征聚合成维度二, 见表3, 其中3项负向特征均属动词词类。“副动词”和“形式动词”用以增强句中谓语的陈述性和宾语的指称性, 构成以合偶词为基础的“双必合双”结构(冯胜利等, 2008; 王永娜, 2016), 如“设法完成”“加以警告”等, 呈现“对行为的再现或重构”功能, 一般用于说明类和实证类学术论文。

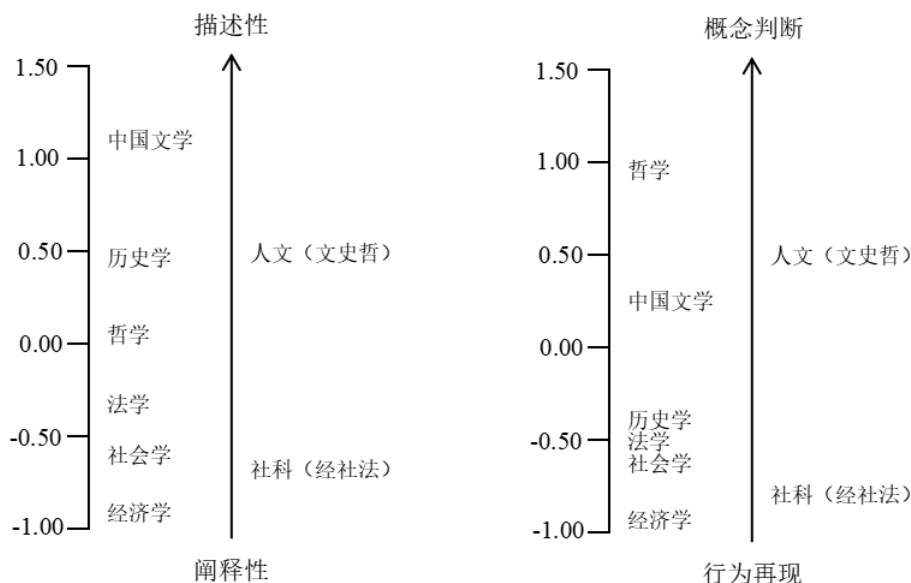


图 1: 维度分均值在维度一和维度二的分布

维度	语言特征	载荷值	语言特征	载荷值
维度二	第三人称代词	0.852	心理名词	0.359
	动词“是”作主要动词	0.744	推测性动词	0.339
	介词“把”	0.744	趋向动词	0.319
	代词：它	0.691	插入语	0.315
	助词“的”	0.570	形式动词	-0.309
	数量词	0.473	副动词	-0.310
	大学学科专业分类词	0.392	动词：非常用	-0.368
	比况助词	0.387		

表 3: 维度二的语言特征及其载荷值

维度二的正向特征“第三人称代词”“动词‘是’作主要动词”“代词‘它’”等，在书面语中用于对客观的人、事物或现象进行概念判断或特点概括，是一种“下定义”式的表达，以突出其固有的、区别于他者的特点。维度二的其他正向特征，如“数量词”、“大学学科专业分类词”、“比况助词”等起到限定、指代、说明等作用，让一些抽象概念被解析得更加明确。

语料 (3) 节选自哲学期刊论文，标注出维度二的部分正向特征（代词：它、动词“是”作主要动词、介词“把”、助词“的”）。语料 (4) 节选自社会学期刊论文，标注出维度二的负向特征（动词：非常用、副动词、形式动词）。

(3) 第二种理论主张可称为有意识的倾向主义。它坚持认为，一个命题态度是由某些对意识显现状态倾向所构成的。第三种可称为自然种类主义。它坚持认为，一种命题态度是无论什么自然种类发挥的所有作用，既是行为的又是现象的，并且我们把这种作用相关于这种态度。

(4) 在使用年龄偏好指数进行相应计算时有一个前提假设条件。同时，单岁组计算时分母相应增多，也稀释了流动人口集中带来的影响。对此，本文提出对联合国综合指数进行修正，以单岁组代替5岁组为单位进行分析计算。

从图1的维度分均值分布可知，中国文学和哲学在维度二均值为正，尤以哲学表现出较强的“概念判断”特征；而历史学和社科领域的三门学科为负值，表现出与之对照的“行为再现”特征。从某种意义上来说，历史学也是基于史料的论证和探索，包含对先前研究行为的再现，故一些历史学论文在该维度上呈现出“行为再现”的功能。这体现了历史学在表达“行为再现”功能时区别于其他人文学科的独特性。

### 3.3 维度三：铺陈与发展

维度三的语言特征包括四类复句和两种副词，均为正向，且绝对值均大于0.4，见表4。在汉语中，复句各分句之间关系复杂，但表述层次清楚，语意周密(吴礼权, 2012)。该维度的复句中，三种属于偏正复句，一种（顺承复句）属于联合复句，使成分之间联系紧密，铺陈接续。

维度	语言特征	载荷值	语言特征	载荷值
维度三	因果复句	0.771	时间副词	0.498
	条件复句	0.715	副词：表态度	0.474
	顺承复句	0.626	假设复句	0.432

表 4: 维度三的语言特征及其载荷值

哲学期刊论文在该维度的均值最高（图2），节选语料（5）标注出两种复句（因果复句、条件复句），体现出较强的逻辑关系。

(5) 在基督教中，无论是否亲历基督复活之事，一个人只要相信此事，他就在精神上成为了圣保罗的当代人，……之所以是复活而不是回忆，因为历史虽可以回想，却无法在苏格拉底意义上被回忆。……只有开天辟地的造物主才能在无宾语状态下仅凭自身而存在。

基于以上特征共同表达的功能，将维度三命名为“铺陈与发展”。这是一种对客观事实的描述和看法，强调时间或逻辑上的紧密联系，包含一定程度的主观性与非理性因素，具有人为参与感。

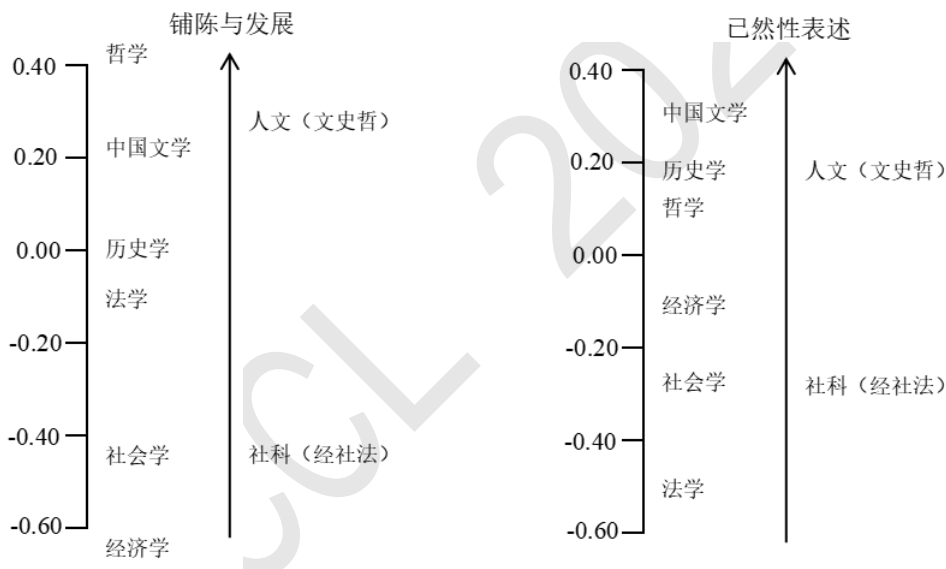


图 2: 维度分均值在维度三和维度四的分布

### 3.4 维度四：已然性表述

维度四由“过去式动词”和“助词‘了’”组成，均具有极高载荷值，见表5。二者关联性较强，都是对过去发生的动作或行为的表述。

维度	语言特征	载荷值	语言特征	载荷值
维度四	过去式动词	1.030	助词“了”	0.988

表 5: 维度四的语言特征及其载荷值

以汉语为代表的孤立语主要通过虚词的语法手段来表达时间范畴。本文中，“过去式动词”的检索方式为“表示过去的副词+动词”，较多分布在陈述性的话语中。助词“了”是汉语的三

个动态助词之一，除了表示过去的意义，还可表达完成的意义。由此可见，维度四的两项语言特征表达相似的时和体意义，具有高度相关的功能。

语料（6）节选自中国文学论文，标注了维度4的语言特征（助词“了”、过去式动词），更加清晰地看出作者对已然性的表述。文中通过对莫言人生经历的回顾，表现作者对莫言及其作品研究的历程，为文章增添说服力。

（6）理论上讲，作家莫言早已成了都市人。……近年来，有诸多研究者已发现莫言与沈从文经历的相近，……这一写作模式深刻影响了几代中国作家的回乡写作。……事实上，批评家们已然注意到，以《红高粱》为标志，莫言已经寻找到他的“精神家园”……她嫁了个哑巴丈夫，生了个哑巴孩子。

由图2可知，人文领域的三门学科在该维度均为正值，而社科领域的三门学科均为负值，呈现出与维度一类似的学科分野。事后检验结果表明，该维度中人文领域的三门学科之间均不具有显著性差异，说明“已然性表述”亦是人文领域的常见表达功能。

### 3.5 维度五：计数与测量

维度五共4项特征，均为数词和量词，见表6。在汉语中，二者通常组合成数量短语作句法成分，因而其出现频率和分布具有正相关。量词具有一定的感情色彩义，对于相同的对象，使用不同量词会产生不同的感情色彩。由于科学语体在词语运用上排斥描绘形象色彩和感情色彩等(袁晖, 李熙宗, 2005)，因而法学期刊论文中数词和量词出现频率较低，见图3。同理，哲学概念往往由于其抽象性而无对应的量词，故哲学语域表现出与法学类似的情形。

维度	语言特征	载荷值	语言特征	载荷值
维度五	动量词	-0.315	数词	-0.822
	时量词	-0.678	量词	-0.867

表 6: 维度五的语言特征及其载荷值

语料（7）和（8）分别来自经济学和历史学，标注了维度五的全部四项语言特征（量词、数词、时量词、动量词），从中可见对“计数与测量”表达的严密性和客观性。

（7）利用中国85个大中城市的环境污染和房价数据，……这是首次运用发展中国家的数据检验城市化进程中……近10年来，中国经历了快速的房价上涨，……这4个城市的房价分别上涨了2.18倍、1.89倍、2.16倍和1.81倍。

（8）每在黄道上移动15度时，即定为一节气的日期。……以致有可能出现一月之中有三气，偶而也会出现一年之内有两个无中气之月的现象。……清初以来三百五十年间，平均每半个世纪才发生一次的闰八月，……将是社会天文学史研究的另一重要内容。

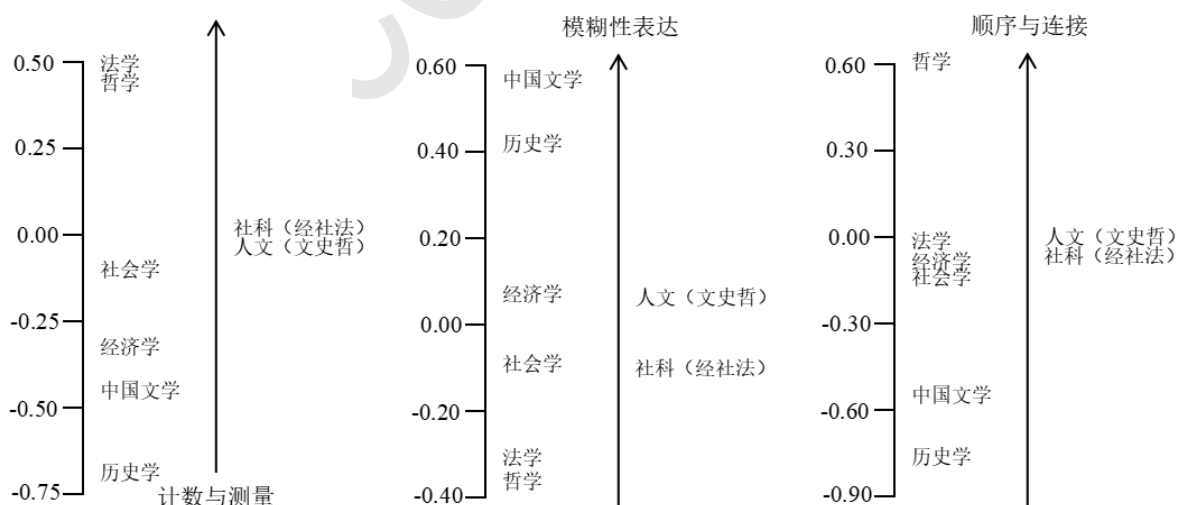


图 3: 维度分均值在维度五、维度六和维度七的分布

### 3.6 维度六：模糊性表达

维度六包含4项语言特征，见表7。其中“模糊限制语”和“副词：表可能”载荷值均大于0.5，二者分别是对模糊性和可能性的表达，且存在一定重叠，在使用中具有较高关联性和替代性。“转折复句”通过关联词语表达转折关系范畴，在语义特征方面具有相对性和补充性，体现信息的翻转和婉转(赵岩, 2021)，反映交互双方的信息不确定或不对称。“插入语”属于独立语范畴，用于补足句义，使句子表意更加严密，如“据说”“在我看来”等。

维度	语言特征	载荷值	语言特征	载荷值
维度六	模糊限制语	0.868	转折复句	0.491
	副词：表可能	0.643	插入语	0.350

表 7: 维度六的语言特征及其载荷值

以上特征共同实现对事物发展变化的模糊性推断和判断，表示不肯定或不确切的意味。由图3，中国文学和历史学在该维度体现明显。语料(9)节选自历史学期刊论文，标注维度六的全部特征(模糊限制语、副词：表可能、转折复句、插入语)，表达因年代久远而对历史事件的推测或假设。

(9) 这当然只是胡适和陈垣个人之间的争论，其中或许<sup>1</sup>有胡适和陈垣之间的瑜亮心结，所以话题变得比较敏感。但是，在现代中国的佛教道教研究史中……不过，事实上彼此也还是有些在意。4月10日，陈垣写了一封似乎<sup>2</sup>是高挂“免战牌”的信，……换句话说，即佛教传到中国，……是否还可能经由南海到达交、广地区这一途径？

### 3.7 维度七：顺序与连接

维度七由载荷值均较高的“顺序词”和“句内并列连词”组成，见表8。在汉语中，二者都可归为连词范畴，都是虚词。顺序词体现先后的逻辑，并列连词是对具有并联关系成分间的连接。该两项特征均属于明晰化手段，能让文本的层次、语义关系变得明白确切(袁晖, 李熙宗, 2005)，增强表达的精确性。

维度	语言特征	载荷值	语言特征	载荷值
维度七	顺序词	1.097	句内并列连词	0.897

表 8: 维度七的语言特征及其载荷值

由图3，哲学论文在该维度的表现尤为突出。语料(10)节选自哲学期刊论文，标注维度七的语言特征(顺序词、句内并列连词)。句中并列连词和顺序词对文本架构进行梳理和推动，使行文在叙述抽象概念时更有逻辑和层次。

(10) 对于确定性的追问首先<sup>1</sup>表现在对世界存在、变化和发展的始基、根据和解答上。……其次<sup>2</sup>，经过实在与表象、本质与现象二元论的剥离，……最后，以同一性排斥差异性，并最终导致……人类认识的起始和真理标准及意义界限等……

## 4 学科语域的差异检验及聚类分析

### 4.1 基于方差分析的差异检验

通过维度分均值对比和语料分析，本文从7个维度对不同学科语域的差异进行分析。为了验证以上差异在哪些方面具有统计学意义的显著性，本文使用多变量方差分析(MANOVA)对各文本的7个维度分数据进行差异检验。

首先，使用基于马氏距离(Mahalanobis distances)的方法对多元异常值进行检测并剔除。结果显示，数据中有17项异常值，剔除后剩余953篇有效语料。同时，使用皮尔逊相关系数对多因变量的多重共线性进行检验<sup>3</sup>，发现各因变量间相关系数的最大值 $r=0.591 < 0.90$ ，即因变量间的多重共线性不强，满足进行后续统计分析所需条件。

<sup>3</sup>Tabachnick and Fidell(2013)指出，具有较强相关性的因变量之间存在多重共线性，其相关系数为 $r=0.90$ 或更高。故使用皮尔逊相关系数对多因变量进行多重共线性检验，若 $r$ 值小于0.90，则可认为因变量之间的多重共线性较弱，对结果的影响较小。



接着，以领域（人文、社科）、人文学科（文、史、哲）和社科学科（经、社、法）分别作为分组变量，对各组内的文本维度分进行三次协方差矩阵齐性假设检验，发现三个Box’s M协方差齐性检验结果均显著。进而采用Pillai’s trace校正统计量<sup>4</sup>分别再次检验，发现三次Pillai’s trace检验的p值均接近0，说明三种分组情况下的文本维度分在本文建立的7个维度上具有显著性差异。

随后，为验证各单一维度下的学科语域差异是否显著，使用单变量方差分析，将单一维度下的误差方差进行分解。方差齐性检验表明，各单一维度下的方差同质性Levene’s检验结果均显著。为此，使用修正方法Welch’s ANOVA对三种分组方式下每个维度的文本维度分数据进行组间差异性检验，领域分组检验结果如表9所示，学科分组主体间效应检验结果见附录B。

来源	因变量	III类平方和	自由度	均方	F值	p值	$\eta^2$
领域	维度一分数	300.253	1	300.253	491.105	0.00	0.341
分组	维度二分数	365.710	1	365.710	656.067	0.00	0.408
	维度三分数	108.221	1	108.221	144.646	0.00	0.132
	维度四分数	55.468	1	55.468	58.004	0.00	0.057
	维度五分数	0.069	1	0.069	0.087	0.768	0.00
	维度六分数	3.808	1	3.808	4.538	0.033	0.005
	维度七分数	0.188	1	0.188	0.196	0.658	0.00

表 9: 人文社科领域7个维度文本维度分的主体间效应检验结果

结果显示，人文和社科领域之间的语域变异显著存在于5个维度（p值小于0.05），分别是：维度一“描述性vs.阐释性”、维度二“概念判断vs.行为再现”、维度三“铺陈与发展”、维度四“已然性表述”和维度六“模糊性表达”。同时，文、史、哲期刊论文或经、社、法期刊论文各有6个维度呈现显著性差异，说明这两大领域内部学科亦存在凸出的语言变异情况。

为了探究人文和社科两大领域内部的学科在7个维度上各自存在哪些差异，本文以上述结果为基础，在95%置信区间内采用Dunnett’s T3检验法<sup>5</sup>进行事后检验，各学科的多重比较显著性结果可概括为表10，详细结果见附录C。

	中国文学	历史学		经济学	社会学
历史学	I,II,III	/	社会学	I,II,V	/
哲学	I,II,V,VI,VII	I,II,III,V,VI,VII	法学	I,II,III,IV,V,VI	I,III,IV,V

表 10: 事后检验呈现显著差异的维度简表

以上结果表明，人文领域内部学科在除维度四“已然性表述”外均具有显著差异，而社科领域内部学科在除维度七“顺序与连接”外均具有显著差异。其中，哲学与中国文学在5个维度存在显著差异，与历史学在6个维度存在显著差异，说明哲学语域跟同属人文领域的中国文学、历史学语域存在较大差异；法学与经济学在6个维度上存在显著差异，与社会学在4个维度上存在显著差异，说明法学语域与同属社科领域的经济学、社会学语域差异较大。

#### 4.2 学科语域的聚类分析

为了进一步量化比较不同学科语域之间的差异，本文使用Minitab软件进行聚类分析。具体方法为：以构成7个维度的60项语言特征的均值为依据，通过计算值间距离，对六门学科语域进行聚类。相关参数设定为：联结法为最长距离法，距离度量为Euclidean平方，标准化变量，相似性水平为95%。聚类结果如图4所示。

<sup>4</sup>对于Pillai’s trace校正统计量，Pillai(1955)指出该方法通常被认为是最有效且稳健的，是在不满足协方差矩阵同质性假设的情况下进行MANOVA的最佳选择。

<sup>5</sup>根据Sauder and DeMars(2019)，在不满足方差齐性的情况下，使用Dunnett’s T3进行事后检验能够有效控制第一类错误，是一种比较保守的多重比较方法。

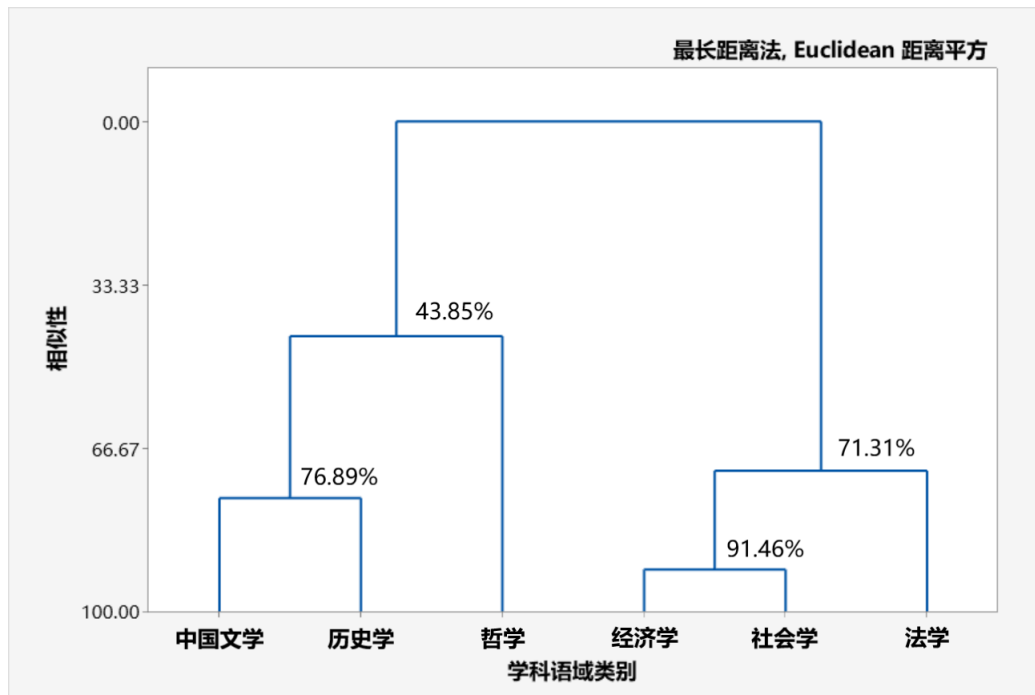


图 4: 学科语域的聚类

由图可知，六门学科的聚类结果分明，共有两个层次。首先，六门学科被分为两大聚类集合：第一大聚类集合包括中国文学、历史学、哲学，第二大聚类集合包括经济学、社会学、法学——这与本文选取的“人文”和“社科”两大领域下的学科分布完全一致。其次，在人文领域内部，中国文学与历史学的相似度为76.89%，哲学与其相似度较低，仅为43.85%；在社科领域内部，经济学与社会学的相似度为91.46%，法学与其相似度为71.31%。以上数据实现了对六门学科之间文本相似性的度量，相关结论与事后检验结果一致，从而量化反映出人文社科领域语域及其内部学科语域之间的差异大小。

## 5 结语及展望

语体语法是现代汉语领域新近形成的重要语法研究流派。该流派秉持“大语法观”，认为语体语法的特征体现于“音系、韵律、词法、句法、篇章、语义”各层面语言表达中(冯胜利, 施春宏, 2018)，凡语言系统中某种结构形式的变化和某项特征的有无或差异能够体现为语体功能的对立或改变，都可成为实现语体的表现手段和方式。

以上观点与本文所使用的语域多维度分析法的底层逻辑存在高度一致性。具体而言，本文发现：汉语学术语体内部存在显著的语言形式与功能差异，仅以其中的人文社科期刊语域而言，其语法范畴层面的语言特征在使用频数上具有明显差异，这些差异经过因子分析可概括为7个语言功能差异维度（即：描述性vs.阐释性、概念判断vs.行为再现、铺陈与发展、已然性表述、计数与测量、模糊性表达、顺序与连接）。并且，人文与社科语域在其中五个维度的表征上存在显著差异，人文或社科内部各自在其中六个维度的表征上存在显著差异。

这为语体语法理论的发展完善提供了具体的资料与例证。尤其是考虑到当前汉语语体语法的研究更多采用讨论某一语法特征在两种特定语体中“合法”或“不合法”的两极判断范式，而以本文所示的研究范式，则有助于构建语法特征在不同的语体/语域中由“不合法”到“合法”的连续统，同时也可以看到多个语法特征使用频数的差异是如何实现不同的差异维度（语言功能），而若干语言功能的定量表达又是怎样构建了不同的语域/语体，这对于丰富和发展汉语语体语法理论有重要启示作用。

此外，本文的发现对人文社科领域的人才培养，特别是学术写作教学实践也具有指导作用。当前国内高校对学术写作的教学指导多停留在学位论文写作方面。诚然，学位论文固然重要，但其仅仅是学术写作诸多形式之一，况且当前学位论文写作的教学重点更多强调学位论文的程序性和规范性，而非学术汉语语言特征的使用。本文通过人文社科中文学术期刊语言差异

研究的发现, 凸显了学术汉语语言表征的多样化功能及形式的丰富性, 在一定程度上揭示出人文和社科领域六个具体学科的学术语言表征特点, 对于学术写作教学有具体的参照借鉴意义。

后续工作可从以下方面继续深入: 1) 在学科语域选择方面, 引入更多领域和学科进行对比, 如人文社科与理工学科的对比, 语言学、心理学等跨领域多栖学科的对比等; 2) 在语料库建设方面, 除了继续扩充学术期刊、学位论文、学术教材等书面语语料外, 亦可增添学术讲座、课堂授课、研讨会发言等学术口语语料, 丰富学术汉语语料库的类型和规模, 从而更加全面地考察学术汉语语域变异情况。3) 另外, 相关应用研究也可开展, 如: 基于学科语域变异规律的国内高校学术写作教学研究、汉语作为第二语言的学术语体教学研究、“中文+职业教育”分科教材编写研究等。

## 参考文献

- Barbara G. Tabachnick, Linda S. Fidell. 2013. *Using multivariate statistics(6th edition)*. Boston: Pearson Education, Inc.
- Derek C. Sauder, Christine E. DeMars. 2019. An updated recommendation for multiple comparisons. *Advances in Methods and Practices in Psychological Science*. 2(1):26-44.
- Douglas Biber. 1985. Investigating macroscopic textual variation through multi-feature/multi-dimensional analysis. *Linguistics*. (2):337-360.
- Douglas Biber. 1986. Spoken and written textual dimensions in English: Resolving the contradictory findings. *Language*. (2):384-414.
- Douglas Biber. 1988. *Variation Across Speech and Writing*. Cambridge: Cambridge University Press.
- Douglas Biber. 1995. *Dimensions of register variation*. Cambridge: Cambridge University Press.
- Douglas Biber. 2006. *University Language: A Corpus-based Study of Spoken and Written Registers*. Amsterdam: John Benjamins Publishing Company.
- Douglas Biber. 2014. Using multi-dimensional analysis to explore cross-linguistic universals of register variation. *Languages in Contrast*. (1):7-34.
- Douglas Biber, Bethany Gray, Shelley Staples. 2016. Predicting patterns of grammatical complexity across language exam task types and proficiency levels. *Applied Linguistics*. 37(5): 639-668.
- Douglas Biber, Jesse Egbert. 2016. Register variation on the searchable web: A multi-dimensional analysis. *English Linguistics*. 44(2):95-137.
- Douglas Biber, Jerry Kurjian. 2006. Towards a taxonomy of web registers and text types: A multidimensional analysis. *Language and Computers*. (1):109-131.
- Douglas Biber, Susan Conrad. 2019. *Register, Genre and Style (2nd edition)*. Cambridge: Cambridge University Press.
- K. C. S. Pillai 1955. Some new test criteria in multivariate analysis. *Annals of Mathematical Statistics*. 26(1):117-121.
- Zheng-sheng Zhang. 2012. A corpus study of variation in written Chinese. *Corpus Linguistics and Linguistic Theory*. 8(1):209-240.
- 刁晏斌. 2004. 试论现代汉语形式动词的功能. 宁夏大学学报. (3):33-38.
- 范楚琳, 刘颖. 2010. 基于多维度分析法的鲁迅三种文体比较研究. 中文信息学报. 34(10):94-104.
- 方梅. 2007. 语体动因对句法的塑造. 修辞学习. (6):1-7.
- 冯胜利, 王洁, 黄梅. 2008. 汉语书面语体庄雅度的自动测量. 语言科学. (2):113-126.
- 冯胜利, 施春宏. 2018. 论语体语法的基本原理、单位层级和语体系统. 世界汉语教学. (3):302-325.
- 胡春雨, 谭金琳. 2020. 中美企业致股东信语域特征的多维分析. 外语与外语教学. (6):66-75.

- 黄梅, 冯胜利. 2009. 嵌偶单音词句法分布刍析——嵌偶单音词最常见于状语探因. 中国语文. (1):32-44.
- 雷秀云, 杨惠中. 2001. 基于语料库的研究方法及MD/MF模型与学术英语语体研究. 当代语言学. (3):143-158.
- 刘艳春. 2019. 汉语语体变异的多维度分析——基于17个语体72项语言特征的考察. 江汉学术. (3):100-110.
- 潘. 2012. 中国非英语专业本科生和研究生书面语体的多特征多维度调查. 外语教学与研究. (2):220-232.
- 王德春. 2004. 大学修辞学. 福州: 福建人民出版社.
- 王永娜. 2015. 汉语合偶双音词. 北京: 北京语言大学出版社.
- 王永娜. 2016. 汉语书面正式语体语法的泛时空化特征研究. 北京: 中国社会科学出版社.
- 武姜生. 2001. 语域变异的多维向分析模式简介. 解放军外国语学院学报. (3):6-9.
- 武姜生. 2004. “学术交流e-mail”文体特征的多维度分析. 外语与外语教学. (2):53-57.
- 吴礼权. 2012. 现代汉语修辞学. 上海: 复旦大学出版社.
- 袁晖, 李熙宗. 2005. 汉语语体概论. 北京: 商务印书馆.
- 张伯江. 2012. 以语法解释为目的的语体研究. 当代修辞学. (6):13-22.
- 赵岩. 2021. 现代汉语转折关系范畴研究. 长春: 吉林大学博士学位论文.
- 朱晓楠. 2014. 汉语普通话的多维语域变异研究. 杭州: 浙江大学硕士学位论文.
- 朱宇, 胡晓丹. 2021. 汉语连词在不同学术语域的聚合: 多维度定量分析. 语言教学与研究. (2):57-69.

## 附录A.现代汉语语言特征体系

- 名词及其特殊属类:** 1.名词: 最常用; 2.名词: 中度常用; 3.名词: 非常用; 4.抽象名词; 5.立场名词; 6.心理名词; 7.大学学科专业分类词; 8.指人名词; 9.集体名词; 10.具象名词; 11.具象科技名词; 12.度量衡名词
- 动词及其特殊属类:** 13.动词: 最常用; 14.动词: 中度常用; 15.动词: 非常用; 16.动作行为动词; 17.使役动词; 18.存现动词; 19.心理动词; 20.肯定动词; 21.交流动词; 22.推测性动词; 23.可能性动词; 24.趋向动词; 25.副动词; 26.形式动词
- 形容词及其特殊属类:** 27.形容词: 最常用; 28.形容词: 中度常用; 29.形容词: 非常用; 30.大小形容词; 31.常用形容词: 相关性; 32.状态词; 33.区别词
- 数词和量词:** 34.数词; 35.数量词; 36.量词; 37.动量词; 38.时量词
- 代词和代动词:** 39.第一人称代词“我”; 40.第一人称代词“我们”; 41.其他第一人称代词; 42.第二人称代词; 43.第三人称代词; 44.代词: 它; 45.普通名词+们; 46.指示代词; 47.不定代词; 48.句首代词; 49.代动词
- 副词及其特殊属类:** 50.副词: 最常用; 51.副词: 中度常用; 52.副词: 非常用; 53.副词: 表必然性; 54.副词: 表可能; 55.副词: 表态度
- 介词及其短语:** 56.句首介词; 57.介词短语
- 助词:** 58.助词“的”; 59.助词“地”; 60.助词“得”; 61.助词“等”/“等等”; 62.比况助词
- 其他词汇属类:** 63.语气词; 64.顺序词; 65.句内并列连词; 66.小品词; 67.口语词; 68.儿化词; 69.嵌偶单音词; 70.合偶双音词; 71.古语词; 72.增强语; 73.模糊限制语
- 名词形式:** 74.名物化; 75.名动词; 76.名形词
- 状态形式:** 77.动词“是”作主要动词; 78.动词“有”, 表存现
- 时态和体标记:** 79.进行式动词; 80.过去式动词; 81.助词“着”; 82.助词“了”; 83.助词“过”
- 情态动词:** 84.必要性情态动词; 85.情态动词: 表未来
- 地点和时间状语:** 86.时间副词; 87.副词: 表处所; 88.处所词; 89.方位词

缩略形式: 90.缩略语  
 否定形式: 91.否定词  
 独立语: 92.叹词; 93.拟声词; 94.插入语  
 疑问句: 95.特指疑问句  
 被动形式: 96.被动; 97.无施事者被动句; 98.介词“把”  
 复句: 99.并列复句; 100.顺承复句; 101.解说复句; 102.选择复句; 103.递进复句; 104.条件复句; 105.假设复句; 106.因果复句; 107.目的复句; 108.转折复句  
 词汇丰度: 109.类符/形符比; 110.词长; 111.平均句长

## 附录B.学科分组主体间效应检验结果

文史哲三学科7个维度文本维度分的主体间效应检验结果

来源	因变量	III类平方和	自由度	均方	F值	p值	$\eta^2$
人文 分组	维度一分数	122.943	2	61.471	87.686	0.00	0.245
	维度二分数	135.431	2	67.761	137.736	0.00	0.338
	维度三分数	15.489	2	7.744	8.979	0.00	0.032
	维度四分数	3.645	2	1.823	1.582	0.206	0.006
	维度五分数	138.147	2	69.074	96.284	0.00	0.263
	维度六分数	111.104	2	55.552	66.571	0.00	0.198
	维度七分数	211.572	2	105.786	119.934	0.00	0.308

经社法三学科7个维度文本维度分的主体间效应检验结果

来源	因变量	III类平方和	自由度	均方	F值	p值	$\eta^2$
社科 分组	维度一分数	15.041	2	7.520	47.179	0.00	0.188
	维度二分数	11.418	2	5.709	19.728	0.00	0.008
	维度三分数	21.096	2	10.548	20.521	0.00	0.092
	维度四分数	11.395	2	5.698	8.516	0.00	0.040
	维度五分数	49.154	2	24.577	56.753	0.00	0.218
	维度六分数	11.403	2	5.701	10.314	0.00	0.048
	维度七分数	0.066	2	0.033	0.06	0.942	0.00

附录C.人文社科领域内部学科在7个维度的事后多重比较显著性结果 ( $p$ 值)

维度一：描述性vs.阐释性					
	中国文学	历史学		经济学	社会学
历史学	0.000	/	社会学	0.000	/
哲学	0.000	0.000	法学	0.000	0.021
维度二：概念判断vs.行为再现					
	中国文学	历史学		经济学	社会学
历史学	0.000	/	社会学	0.008	/
哲学	0.000	0.000	法学	0.000	0.368
维度三：铺陈与发展					
	中国文学	历史学		经济学	社会学
历史学	0.037	/	社会学	0.074	/
哲学	0.106	0.000	法学	0.000	0.016
维度四：已然性表述					
	中国文学	历史学		经济学	社会学
历史学	0.521	/	社会学	0.543	/
哲学	0.238	1.000	法学	0.000	0.039
维度五：计数与测量					
	中国文学	历史学		经济学	社会学
历史学	0.214	/	社会学	0.009	/
哲学	0.000	0.000	法学	0.000	0.000
维度六：模糊性表达					
	中国文学	历史学		经济学	社会学
历史学	0.282	/	社会学	0.310	/
哲学	0.000	0.000	法学	0.000	0.063
维度七：顺序与连接					
	中国文学	历史学		经济学	社会学
历史学	0.251	/	社会学	0.999	/
哲学	0.000	0.000	法学	0.993	0.984