

基于文本分析的来华留学生对中国的情感认知研究 ——以“南亚师资班”作文为例^{*}

张黎^{*} 陈光辉

摘要

以 148 篇与中国相关的南亚师资班作文为语料，利用智能化文本分析软件提取了十个有代表性的中国元素主题词作为情感分析的测量变量。通过基本情感分析与交叉分析得出了一系列情感认知数据，并通过检索返回语料验证了数据的有效性。研究发现南亚留学在“中国”这一测量变量上表现出正面情感较强，内部一致性较好等特点，此外也发现学习上的困难体验对其情感认知的负面影响较大，留学所在城市对其情感认知具有较大影响，留学时间长短与情感倾向呈反向发展态势（有待进一步验证）。研究所用方法可作为传统问卷调查方法相互补充，研究成果对于全面认识来华留学生的情感认知特点及如何更好地举办南亚班这种政府留学项目可以提供启示和借鉴。最后指出了研究有待完善之处。

关键词：南亚师资班 来华留学生 情感认知 文本情感分析

^{*} 张黎，北京语言大学教师教育学院，联络电邮：jmx01@blcu.edu.cn。（本文通讯作者）

陈光辉，北京十一学校天津实验学校，联络电邮：2902020362@qq.com。

^{*} 本研究得到 2020 年度国家社科基金重大项目资助（项目批准号：20&ZD311）。

引言

留学生在留学目的国的学习与生活体验，会直接影响其对该国的认知，并形成其对该国事物的主观情感，这就是留学体验产生的情感认知。我国来华留学生教育，特别是对外汉语教学界，比较重视培养留学生对中国的情感，学界也有一些研究。以往研究主要运用问卷调查的方法来获取来华留学生的情感信息，这种方法虽然有效，但有其局限：一是评价要素及角度人为设定，未必全面反映被试的真正关注点，也容易遗漏一些可能是更重要的方面；二是研究意图和目的有预设，并为被试察觉和意识到，影响其客观性。那么有没有相对客观的方法来研究这一问题呢？我们在对孔子学院总部/国家汉办（以下简称孔院总部）主持的“南亚国家汉语师资班”政府奖学金项目（以下简称“南亚班”）的学生进行写作教学的过程中，发现其作文文本能够反映出其对中国的情感认知的一些“蛛丝马迹”，由此我们就想到可以尝试对其作文进行文本分析来发现其对中国情感认知的特点。因此我们考察了现有的一些文本分析软件，最后选择了“内容挖掘系统”（ROST CM 6.0）和“自然语言处理系统”（NLPIR）作为研究工具，以南亚班作为案例，以其与中国相关的命题作文文本语料为分析样本作一研究探索。研究目的主要是客观揭示南亚班留学生在留学的过程中，对有关中国的哪些元素有更深体会，对这些元素的情感如何，不同性别、年级、国别的留学生之间是否有显著差异。同时，也探索通过文本内容挖掘分析，研究来华留学生情感认知的方法和路径，尝试突破现有研究的局限。该研究成果对于全面认识来华留学生的情感认知特点及如何更好地举办南亚班这种政府留学项目可以提供启示和借鉴。

一、相关研究回顾

来华留学生对中国的情感认知的研究，近十几年才得到关注，并不断有研究成果发表。现从以下几方面作一回顾。

（一）主要研究内容

对中国文化和国家形象认知的调查研究。部分学者依据现有的理论将文化分为不同的类型，在此基础上研究来华留学生各种类型的中国文化或某一类中国文化的认知问题。如：王强（2015）从物质、观念和行为三个层面调查留学生对中国文化的认知情况；高金萍、王纪澍（2017）调查了北京地区五所高校来华留学生对中国物质文化和精神文化的认知情况；哈嘉莹、尚晓燕（2017）调查了“一带一路”沿线国家来华留学生对中国物质文化形象的认知情况；邵倩、郑志刚（2010）调查了外国留学生对中国传统体育文化的认知情况。王祖嫒（2018）从汉语国际传播的角度出发，探讨了东南亚五国汉语国际传播与中国国家形象认知的相关性；牟蕾等（2019）以留学生心目中的中国形象为主题进行了词汇联想测试。

（二）主要研究对象

在具体考察的对象或变量指标上，不同研究关注和选择的中国元素不尽相同。王强（2015）选择的元素包括传统游戏、集体价值、宴席规矩、传统节日、午休习惯等；高金萍、王纪澍（2017）将汉语、长城、书法、中医、儒家思想等作为物质文化和精神文化的代表；哈嘉莹、尚晓燕（2017）的物质文化要素包括饮食、景点、城市、人物、艺术、运动、医药、服饰、颜色、节日、动物 11 种；而牟蕾等（2019）则将中国形象归为九大类：文化形象、宏观形象、地域形象、饮食形象、人物形象、生活形象、经济形象、政治形象、环境形象。由此可见，同样是调查来华留学生对中国的情感认知，但各家的分类标准、角度不同，考察的中国元素也不尽一致。

（三）主要研究结论

以对中国文化和国家形象的肯定、赞赏态度为主。如邵倩、郑志刚（2010），王强（2015），哈嘉莹、尚晓燕（2017），高金萍、王纪澍（2017），王祖嫒（2018）等等。只有牟蕾等（2019）在考察留学

生评价中国的词语的感情色彩时得出其情感以中性为主（占 89%）的结论。

（四）主要研究方法

以问卷调查和访谈方法为主。如邵倩、郑志刚（2010），王倩（2014），哈嘉莹、尚晓燕（2017），熊智伟（2018），朱佳妮、姚尹喜（2019），吴晶、王睿（2019）等，分析方法也是以定量分析为主。王祖嫒（2018）以宏观系统分析法和全球视野比较法为方法论指导，综合运用了抽样调查、内容分析、田野调查、叙事研究等方法。牟蕾等（2019）采用词语联想测试的方法，让被试对“中国”这一概念自由联想，然后将联想到的词语按照语义的相似度进行归类，计算各类的频次及占比，并对这些词语按其感情色彩进行分类统计，以分析被试的情感倾向。

本研究与已有研究的不同主要表现在以下三方面：

1. 研究范围不同。以往的研究虽然也有针对某一国家或某一区域的留学生，但尚未见到专门针对南亚地区的研究。本研究只针对南亚地区的一个特殊项目的留学生。

2. 材料来源不同。本研究完全依据来华留学生的作文文本，该文本非为研究目的而设计和产出，所以没有预设的中国元素检测指标变量，完全根据文本的自然情况进行分析，因此完全排除了研究者对被试的主观影响。

3. 研究方法不同。本研究主要运用智能化文本分析工具进行情感倾向分析，以计算机自动分析为主，其情感倾向是基于整体文本的综合分析，而非单纯从某个词汇出发。所得情感变量积极或消极倾向以具体的百分比数字来表示，而不是简单分为积极、中性、消极这种属性类别。这不同于王祖嫒（2018）以质性分析为主的叙事研究的方法，也不同于牟蕾等（2019）单纯依据各个词汇的色彩进行分析的方法。

二、研究设计

（一）研究目标

本文的主要研究目标有三：发现南亚班留学生作文文本中最具代表性的中国话题元素；了解南亚班留学生对这些中国元素的基本情感倾向和不同性质留学生之间情感认知的差异；探索文本挖掘技术用于来华留学生情感认知分析的方法和路径。

（二）研究方法及研究工具

本研究采用文本分析的方法，利用智能化文本分析工具对南亚班的作文语料进行纯量化分析。

文本分析是“介于定性和定量分析之间的一类分析方法。其赖以存在的基本假设是：文本是记录我们的思想、感情、观点和信仰的主要方法”（曾忠禄、马尔丹，2011，页48），原理是“通过借助一定的技术手段，将定性的文本材料转化为定量的数据，并依据这些数据对文献内容进行定量化分析，作出关于事实的推断”（毕赛云等，2019，页120），主要技术手段包括词频统计、高频特征词分析、共现网络分析、文本情感分析等。文本情感分析“是对带有情感色彩的主观性文本进行分析、处理、归纳和推理的过程”（赵妍妍等，2010，页1835）。

本研究以武汉大学沈阳研发的内容挖掘系统 ROST CM 6.0 版进行词频统计和共现网络分析，以北京理工大学张华平研发的自然语言处理系统 NLPPIR（即 ICTCLAS 3.0 版）进行情感分析。使用 SPSS 20 对不同性质留学生情感认知的差异进行显著性检验。

（三）研究步骤

1. 语料处理。将原始作文语料以纯文本格式录入电脑，分类并做必要标注。

2. 确定有代表性的中国元素作为测量变量，表现为十个主题词。

具体做法如下：首先借助 ROST CM 6.0 软件对作文文本语料进行分词和词频统计，再利用该软件的共现网络分析功能，参考已有研究，找出十个有代表性的中国元素的主题词作为测量变量，即情感分析的对象。

3. 基本情感分析。利用 NLPPIR 以十个作为测量变量的中国元素主题词进行情感分析，获得情感数据，并展开分析。

4. 交叉分析。将作文文本按照作者的性别、年级、国籍等社会特征变量（自变量）分类，对社会特征与情感倾向进行交叉分析，以发现南亚班学生对中国的情感认知的内部分布差异，并进行差异的显著性检验。

5. 总结结论，讨论相关问题。

（四）语料来源及处理

本研究语料来自“南亚国家汉语师资班”项目学生的命题作文。该项目由孔子学院总部/国家汉办于 2015 年设立，专门面向南亚国家非中国籍人士招生，提供奖学金资助他们到中国高校接受学历教育，旨在培养优秀的本土化汉语教师。主要类型是“一学年汉语强化学习+四学年汉语国际教育本科”型（简称“1+4”）。北语于 2015、2016 年分别接收了 100 余名该项目学生。本研究的语料来自于这两届学生当中已经进入北语本科汉语国际教育专业学习的学生，他们在进入专业学习之前均已通过 HSK5 级考试，所以每个年级汉语水平的起点基本一致。

2016 和 2017 年，孔院总部面向南亚班各组织了一次主题征文活动。2016 年的主题有三个：“‘一带一路’留学故事”（代码 ydyl）、“我和孔子学院的故事”（代码 kygs）、“我的中国梦”（代码 zgm）；2017 年是两个重复的主题“我和孔子学院的故事”（代码 kygs）和“我的中国梦”（代码 zgm）。因为这些主题都直接涉及到中国的话题，所以适合用来分析作者对中国的情感。另外，这两届学生的平时作业中有三个命题作文也涉及中国主题，2015 级有两个，分别是“我对老师

的意见”(代码 yjdc)和“我的中医文化体验之旅”,2016级有一个“旅行感受”(代码 lxgs),我们也将其作为分析语料。语料的作者人数为78个,文本数共148篇,总字数约11万字。各主题文本数及占比见图1。从作者年级分布上看,2015级语料103篇,占70%;2016级45篇,占30%。从性别分布上看,男性作者语料89篇,占60%;女性作者语料50篇,占34%;性别不详的语料9篇,占6%。因有一个作文的主题是“对老师的意见”,有少数同学的作文匿名,所以这部分语料的作者国籍和性别不详。在国籍分布上,语料作者的国籍有巴基斯坦、尼泊尔、孟加拉、马尔代夫、斯里兰卡五个国家,各个国籍作者的语料文本数见图2。

图1:语料的主体分布(单位:篇)

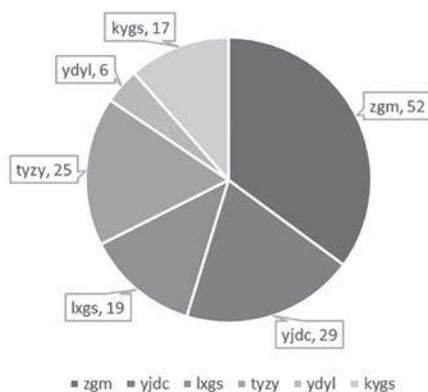
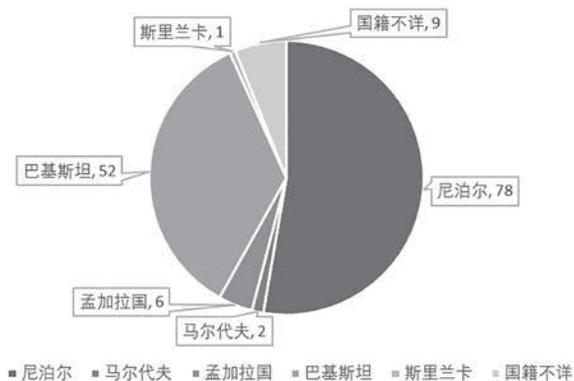


图2:按作者国籍划分的语料分布(单位:篇)

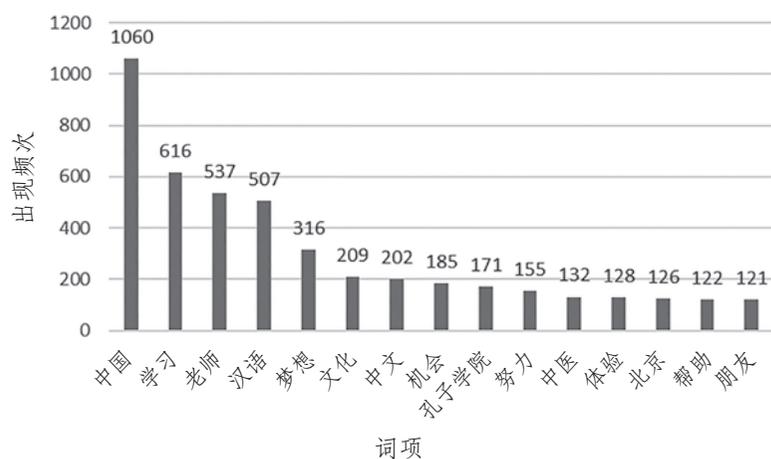


上图还不易观察具体的共现频次，我们又利用 ROST CM 6.0 同时提供的相匹配的高频共现统计表，选取前 15 个与“中国”高频共现的词语，得出高频共现情况和前 15 个高频词（见表 1 和图 4）。

表 1：与“中国”高频近距离共现表

	中国	学习	老师	汉语	梦想	文化	中文	机会	孔子学院	努力	中医	体验	北京	帮助	朋友
中国		289	206	238	155	199	139	148	105	112	49		82	122	103
学习	289		201	222	116	148	126	130	94	108			66	102	82
老师	206	201		152	74	107	74	104	72	92				83	60
汉语	238	222	152		97	125	110	103	81	91				90	74
梦想	155	116	74	97		72	66	60		61					
文化	199	148	107	125	72		70	62	55	54				62	60
中文	139	126	74	110	66	70		67	52						
机会	148	130	104	103	60	62	67		78	53				62	
孔子学院	105	94	72	81		55	52	78							
努力	112	108	92	91	61	54		53							
中医	49											54			
体验											54				
北京	82	66													
帮助	122	102	83	90		62		62							48
朋友	103	82	60	74		60								48	

图 4：前 15 位高频词



可以发现，与“中国”高频近距离共现的14个词正好也是除“中国”外的前14个高频词，经过人工干预之后，我们提出十个主题词作为测量变量，按出现频次高低排序依次为：中国、汉语、学习、老师、梦想、文化、北京、机会、孔子学院、中医。其中“梦想、机会”看似与中国元素无关，但实际上作文的主题有“我的中国梦”与“我和孔子学院的故事”，不少作者把来中国留学看作是一个实现梦想的机会，所以这两个主题词可以视为代表性较强的中国元素。如：

(1) 孔子学院给我机会来中国学习汉语，还给了我当汉语老师的机会，而我也不能错过这个机会。(zgm1525)

(2) 哥哥回答我，在伊斯兰堡的孔子学院，中国政府马上给我们提供学习中文的机会，我完全同意了。(kygs1506)

对中国的情感认知与“机会”和“梦想”相关联，这也是南亚班与其他人所调查对象的明显不同之处。ROST CM 6.0 软件通过语义网络分析自动识别出这两个词与“中国”的搭配关系。所以这十个词，可以反映文本作者对中国的情感体验的来源。下面的基本情感分析和内部差异分析，就主要围绕着这十个测量变量进行。

四、基本情感分析

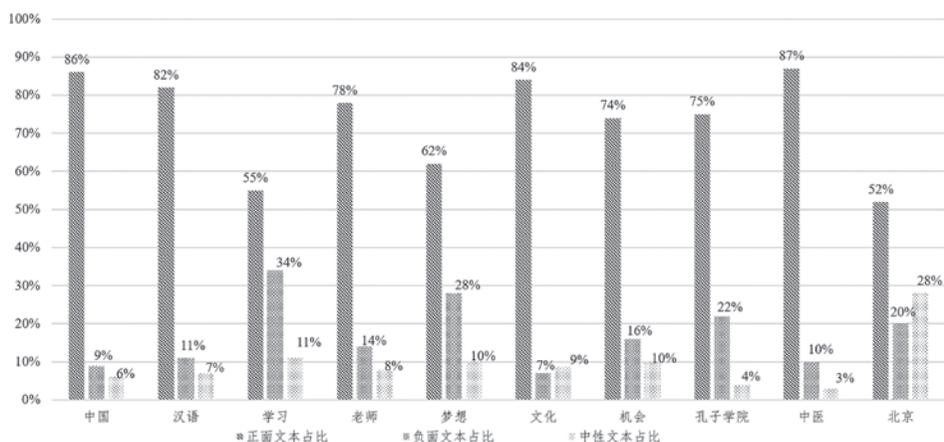
NLPIR 的情感分析功能可以针对事先指定的分析对象，自动分析海量文档的情感倾向，包括情感极性和情感值测量，并在原文中给出正负面的得分和句子样例。例如以“中国”为单一对象进行情感分析，软件可以在148篇文本中自动检索所有包含“中国”的句子并进行情感打分，然后给出每一篇文章的正面得分、负面得分和情感总分，并给出148篇文本中情感总分为正（即正面情感）和情感总分为负（即负面情感）的文本数及对应占比。

基本情感分析就是指对每个测量变量的正面文本占比、负面文本占比和中性文本占比进行综合分析的过程。为验证软件自动分析的可靠性，我们还将以选定的测量变量为关键词通过语料检索返回文本寻找语料支撑，对情感分析的结果进行验证和解释。

（一）基本情感分析数据

从数据看，十项测量变量在 148 篇作文中均以正面形象为主。正面情感文本占比由高到低依次是：中医、中国、文化、汉语、老师、孔子学院、机会、梦想、学习、北京（见图 5）。

图 5：基本情感分析数据



作文的背后是作者，对作文情感的分析结果也反映了作者的情感倾向。在十个测量变量中，“中医、中国、文化、汉语”的正面情感比例在 80% 以上。而“老师、孔子学院、机会”三个变量的正面情感比例在 70%—80% 之间，也算是比较高。而“梦想”的正面情感在 60% 多点，“学习、北京”两个变量的正面情感在 50%—60% 之间，同时，负面情感也相对较高，这也就是说留学生在这几项上的情感总体上一一般。

（二）基本情感分析结果与语料原文的比较

为检验机器分析的可信度，我们通过语料检索的方式进行验证。

语料中谈到“中国”时，往往会联想到发达、历史悠久、风景美丽等（见例 3-4）；提及“汉语”，会与机会和工作联系在一起（见例 5-6）；说到老师和孔子学院，会表示赞赏，有些还明显带有感激之情（见例 7-9）；提到北京、北京人或北京语言大学等与北京相关的

要素，会表达正面的看法（见例 10-11）；谈到中医，会说中医历史悠久、有疗效等（见例 12-13）。

- (3) 我到了中国后发现中国比我想象还发达。(zgm1505)
 - (4) 中国地域广阔，有悠久的历史，有 56 个民族的风俗、文化等等，有特别多美丽的地方。(zgm1514)
 - (5) 而更令我高兴的是，我获得了孔子学院的奖学金，获得了去中国学习汉语的机会。(kygs1502)
 - (6) 到公司以后我才知道，他们需要的翻译不是会英语的，而是要求会汉语的。(kygs1507)
 - (7) 每天有两位老师讲课，他们讲课讲得特别好。(kygs1601)
 - (8) 刚开始学习汉语的时候，我觉得很难，但是我的老师们一直鼓励我，给我加油。(kygs1501)
 - (9) 梦想曾经离我很远，但感谢孔子学院，感谢敬爱的老师们，感谢所有善良的人，现在我正坚定地朝着梦想前进。(kygs1505)
 - (10) 北京是一个古老、时尚的都市。(kygs1504)
 - (11) 北京人很好客，我在北京生活得很开心。(kygs1605)
 - (12) 中医是一门历史悠久的科学，有很多好处。(lxgs1606)
 - (13) 我体验了中医独特的治疗方法，也觉得非常好。(lxgs1606)
- 当然，在谈到某些主题元素时，也会有一定负面情绪的表达。如谈到“学习”，在开始学习汉语的时候会有一定的担忧和畏难情绪（见例 14-15），谈到“中医”，一开始会觉得与西医相比，中医疗效慢、价格高等（见例 16-17），这些都对应机器分析的负面情感。
- (14) 因为中国的学习要求会很严格，孔子学院的学习要求会很高，我能不能合格？(kygs1501)
 - (15) 刚开始学习中文的时候我遇到了很多困难，我一点中文都不会说，老师讲的我也不明白，我的发音也不对，最难的部分是写汉字，我记不住也写不出来。(kygs1503)
 - (16) 因为中医的效果比较好而且很快所以中医比西药贵多了。

(tyzy1523)

(17) 本来我也不是很看好中医，因为我认为中医的治疗效果很慢。(tyzy1525)

从文本的具体内容可以看出，利用计算机所作的情感认知倾向分析，能够真实反映文本中实际表现出来的情感。

五、交叉分析

理论上，不同性别的人的关注点以及对不同元素的敏感程度不同，由此产生的情感倾向也可能不同。年级的不同，意味着来中国留学的时间长短不同，那么其对中国的认知的深度以及情感倾向可能也有所不同。同样，国籍的不同，背景文化不同，对中国的关注点和情感感受可能也不同。因此，我们以作文作者的性别、年级、国别信息作为自变量，来与十个测量变量进行交叉分析，以发现这些自变量在情感认知上是否有分布差异。由于南亚留学生的作文文本在不同测量变量的情感得分不是正态分布，因此，我们采用了非参数检验方法中的曼·惠特尼 U 检验（秩和检验）对不同年级（2015 级和 2016 级）、性别、国别²（巴基斯坦和尼泊尔）的语料在十个测量变量上的情感得分逐一进行了显著性检验，检验工具为 SPSS 20。

结果显示，不同性别、年级、国别的作者对测量变量体现出了比较好的内部一致性。其中，不同性别的作者在十个测量变量的情感得分上均无显著差异。而有显著差异的变量是“汉语、中国、老师”（见表 2）。进一步通过平均秩³分析，可以发现，在变量“汉语”上，2016 级的情感得分（平均秩 = 52.2）总体高于 2015 级（平均秩 = 39.68）；在“中国”上，巴基斯坦作者的情感得分（平均秩 = 63.85）略高于尼泊尔作者（平均秩 = 50.14）；在“老师”上，巴基斯坦作者的情感得分（平均秩 = 60.12）也高于尼泊尔（平均秩 = 47.66）。这也

2 需要说明的是，由于孟加拉国学生的文本数只有六篇，马尔代夫和斯里兰卡的文本数分别只有两篇和一篇，文本数量太少，统计结果不足以说明特征，我们只选择了尼泊尔和巴基斯坦两个国家作者的语料做显著性检验。

3 平均秩：指所有数据按照从小到大排序并编号之后得出的平均值。

就是说，入学时间短的学生对“汉语”的正面情感多于入学时间长的学生，巴基斯坦学生对“中国”和“老师”的正面情感多于尼泊尔学生。对“汉语”的情感年级差异，是随学习时间增长而出现的学习懈怠的表现，从教学实践中可以感受到这种变化，学习时间越长，对汉语的学习兴趣越低。而不同国籍的学生在“中国”和“老师”上的情感差异，目前无法解释，有待进一步探究。

表 2：对十个测量变量情感得分的显著性检验结果

	中国	汉语	学习	老师	梦想	文化	机会	孔子学院	中医 ⁴	北京
分性别	.712	.224	.847	.673	.679	.863	.929	.748	.952	.080
分年级	.127	*.027	.705	.292	.679	.077	.094	.422	.781	.719
分国别	*.028	.221	.601	*.037	.862	.056	.250	.152	.159	.650
Sig > 0.05 无显著意义 Sig < 0.05 有显著意义 Sig < 0.01 有非常显著意义										

六、结论与讨论

通过以上的文本分析，可以看出南亚班学生对中国的情感的一些基本特点：

（一）留学生表现出对中国很强的正面情感认知倾向

148 篇作文中，情感分析结果显示测量变量“中国”正面情感比例高达 86%，可见绝大多数文本均呈现出了很强的正面情感倾向，对其他变量的情感分析也呈现了基本一致的结果。这说明南亚留学生总体上还是非常喜爱中国的。其背后的原因与该项目由中国政府全额奖学金资助的项目性质有关。研究中自动提取出的测量变量包括“机会、梦想”这两个表面上并不直接反映中国元素的词汇项目，表明来中国留学对他们来说是一种美好的事情，这自然影响到他们对中国以及中国文化的强烈好感。同时，关于“孔子学院”这个代表该项目提

4 因不同性别、年级、国别与“中医”相关的文本数较少，因此该列下的 Sig 值为准确显著性，其他各列下的 Sig 值为渐近显著性。

供方的变量，正面情感文本占比排在第六位（75%）也是与其“梦想、机会”的实现有关。以上情况也反映出南亚学生在留学之前就对中国抱有正面的情感。

（二）学习体验对情感认知有一定的消极影响

对于“学习”这一测量变量，不论是整体还是在不同自变量的统计上，相比于其他变量，正面情感文本占比都相对较低，而负面情感文本占比又明显较高，这说明他们当中有不少人在中国的学习体验不太好。在语料中以及我们与他们的现实接触中，都可以看出很多人觉得学习负担重，汉语及专业课程太难，挫败感比较强，还有少数人入系后又转院、转专业。可见，其学习体验的负面情感认知主要是来自于学习上面临的困难。“梦想”的负面情感得分较高，也与学习的艰难有关，学习困难降低了对梦想的期盼。

（三）留学所在地对情感认知有较大影响

分析结果显示，南亚班学生对“北京”的正面情感得分相对较低，占比刚刚过半（52%），这主要是因为北京的气候、自然与社会环境等与他们的国家相差太大，让其难以适应。例如他们普遍觉得北京太干燥，冬天太冷，物价太高等，这些影响了其认知体验和情感倾向。

（四）留学时间长短与情感认知呈反向发展倾向

研究结果显示，虽然两个年级学生对中国元素的正面情感得分都排在第一位，但较高年级学生明显低于低年级，而负向情感得分整体上略高于低年级。按照有关跨文化交际理论以及普遍的感觉，留学时间越长，对留学国的喜爱程度越高，但本研究的结果显示情况并非完全如此，而是随着留学时间的增加，正面情感有所减弱，负面情感有所增强。这应与其学习体验和对中国的更全面的认识有关，不过这点还有待进一步验证。

(五) 情感认知具有较高的内部一致性

交叉分析结果显示南亚师资班作文对中国元素的情感认知表现出较高的内部一致性。十项测量变量中，不同年级仅在“汉语”这一元素上表现出了显著差异性，不同国别仅在“中国”和“老师”这两个要素上表现出了显著差异性。其余变量在不同性别、年级、国别的留学生之间均未表现出显著差异性。体现出较高的内部一致性。

以上研究发现也反映了一些需要注意和改进的方面，包括创造更好的学习体验，以及长期保持甚至提高其正面情感的水平等。

七、研究的价值与不足

本研究尝试利用文本分析技术考察来华留学生对中国的情感认知特征，这种方法能够发现一些典型性和显著性的特征，可以为来华留学生教育实践、教育管理以及政策制定等提供依据。同时，也探索出一条依靠客观观察分析外国学生对中国的情感认知状况的方法，特别是利用语义网络提取测量变量的方法是关键，本研究证明该方法可行、有效，对类似问题的研究提供了新的研究方法，有助于改进和完善这类研究。

本研究当然也存在着一些不足和缺憾，主要在两个方面。一是单纯依靠文本分析得出的结论，是否完全反映真实的情况，只能通过对语料原文进行定性的分析来做有限的验证，如果能够结合其他方面的信息或方法进行验证可能会更有说服力。二是对文本特征的提取精准度可能不够。对于汉语文本的语义的表达有时很难完全通过形式化的特征去捕捉，通过高频词提取文本特征对于判定文本主题具有显著作用，但是是否能够由此就能完全判定作者的情感关注对象还有待进一步探讨。

参考文献

- 毕赛云、李兵、韩玉婷和石金莲（2019）：基于文本分析的国家公园亲子游憩体验研究——以武夷山国家公园试点区为例，《北京林业大学学报（社会科学版）》，18(1)，118-123。
- 高金萍、王纪澍（2017）：来华留学生：中国文化对外传播的重要力量——基于北京地区来华留学生对中国文化认知的调查，《对外传播》，9，50-52。
- 哈嘉莹、尚晓燕（2017）：“一带一路”沿线国家来华留学生中国形象认知研究——基于物质文化的调查分析，《对外传播》，7，52-55。
- 牟蕾、吴勇毅和李婷（2019）：“一带一路”国家来华留学生对中国形象的认知及传播，《青年研究》，5，86-93，96。
- 邵倩、郑志刚（2010）：关于留学生对中国传统体育认知度的调查研究——以郑州大学为例，《衡水学院学报》，12(4)，49-52。
- 王倩（2014）：来华留学生对中国文化认知情况和需求分析的研究——以郑州大学西亚斯国际学院的留学生为例，《时代文学（下半月）》，6，208-211。
- 王强（2015）：外国人对中国文化的认知情况：一项基于留学生群体的调查，《西安外国语大学学报》，2，14-17。
- 王祖嫒（2018）：《东南亚五国汉语传播与中国国家形象认知的相关性研究》，北京，中央民族大学博士论文。
- 吴晶、王睿（2019）：在华留学生对中国文化的认知调查研究——以陕西师范大学为例，《新西部》，17，47-49。
- 熊智伟（2018）：东盟留学生跨文化适应问题实证研究——基于南昌高校东盟的调查，《上饶师范学院学报》，38(1)，109-115。
- 曾忠祿、马尔丹（2011）：文本情感分析方法在竞争情报中的运用，《情报理论与实践》，8，47-50。
- 赵妍妍、秦兵和刘挺（2010）：文本情感分析，《软件学报》，21(8)，1834-1848。
- 朱佳妮、姚尹喜（2019）：外籍留学生对“中国文化”认知、态度和评价的实证研究，《当代传播》，1，56-60，65。

Research on the Emotion and Cognition of China among International Students in China Based on the Text Analysis ——Take South Asia Teacher Class Essays as an Example*

ZHANG, Li* CHEN, Guanghui

Abstract

Taking 148 South Asian teacher class compositions related to China as the corpus, ten representative Chinese element subject words are extracted by using intelligent text analysis software as the measurement variables of emotion analysis. A series of affective cognitive data are obtained through basic affective analysis and cross analysis, and the effectiveness of the data is verified by retrieving corpus. The study found that South Asian students showed strong positive emotion and good internal consistency in the measurement variable 'China'. In addition, it also found that the difficult learning experience had a great negative impact on their emotional cognition, the city where they studied abroad had a great impact on their emotional cognition, and the length of study abroad and emotional tendency showed a reverse development trend (to be further verified). The research results can provide enlightenment and reference for a comprehensive understanding of the emotional cognitive characteristics of foreign students in China and how to better hold South Asia class. Finally, it points out two deficiencies to be improved.

Keywords: South Asia teacher class, foreign students in China, emotional cognition, text emotion analysis

* ZHANG, Li, Teacher College, Beijing Language and Culture University, China. (corresponding author)
CHEN, Guanghui, Beijing National Day School · Tianjin, China.

* This study was supported by the major projects of National Social Science Foundation of China in 2020 (project No.: 20&ZD311).