

基于 CSL 学习者认知的情感词汇计量与统计分析*

张易扬¹, 王治敏^{1*}, 吴迪¹, 张璇¹

(1.北京语言大学, 北京市 100083)

摘要: 本文以情感词汇词典为依托, 通过四部小说情感词汇的提取, 对比和分析四部小说用词的情感分类、词性种类、极性和强度。在此基础上研究汉语作为第二语言 (Chinese as a second language, 简称 CSL) 学习者对“接受性词汇”的情感词汇熟悉度测量表现和“产出性词汇”的情感词汇输出表现, 并进行了科勒-拉普假设检验。我们发现, 现代汉语长篇小说在情感词汇的使用上, 并不会因为作者、题材、内容不同而产生较大差异, 文本中 21 类情感词赞扬类和贬责类占总词数的一半。另外, CSL 学习者对频率高的情感词汇熟悉度不够, 他们在产出情感词汇时动词和形容词产出不够丰富, 悲伤类的词语产出较少, 对表达强烈感情的词汇掌握的也不够多。

关键词: 情感词汇词典; CSL 学习者; 二语习得

Measurement and Statistical Analysis of Emotional Vocabulary Based on CSL Learners' Cognition

Yiyang Zhang¹, Zhimin Wang^{1*}, Di Wu¹, Xuan Zhang¹

(1. Beijing Language and Culture University, Beijing 100083, China)

Abstract: This paper compares and analyses the emotional classification, types of parts of speech, polarity and intensity of the words used in the four novels through the extraction of emotional vocabulary. Meanwhile, we put forward measurement performance of Chinese as a second language (CSL) learners' familiarity with the affective vocabulary of "receptive vocabulary" and the output performance of the affective vocabulary of "productive vocabulary", and carries out Kohler-Rapp hypothesis test. Finally, we find that the use of emotional vocabulary in modern Chinese novels does not vary greatly because of the author, subject matter and content. In the 21 types of emotional words, praise and derogation account for half of the total vocabulary. CSL learners are not familiar with high-frequency emotional vocabulary. In productive vocabulary, verbs and adjectives are not rich enough, negative words are produced less, and vocabulary expressing strong feelings is not enough.

Key words: Emotional vocabulary dictionary; CSL learners; Second language acquisition

* **收稿日期:** **定稿日期:**

基金项目: 国家社科基金重大项目 (18ZDA295); 教育部人文社科基金 (16YJA740036); 中央高校基本科研业务费 (18YBT03, 19YCX057)

作者简介: 张易扬 (1994), 男, 硕士研究生, 主要研究方向: 汉语国际教育; 王治敏, 通讯作者, (1972), 女, 博士生导师, 主要研究方向: 计算语言学, 汉语国际教育; 吴迪 (1994), 女, 硕士研究生, 主要研究方向: 汉语国际教育; 张璇 (1997), 女, 管理学学士, 主要研究方向: 信息管理与信息系统。

1 引言

随着人工智能研究的发展,自然语言处理领域对情感分析、情感计算研究逐渐火热起来。汉语情感词汇相关的词表、词典的构建也逐渐发展起来。有清华大学李军中文褒贬义词典^[1]、台湾大学 NTUSD 简体中文情感词典、知网 Hownet 情感词典、大连理工大学的中文情感词汇本体库、湖南工业大学的中文基础情感词典等。情感词汇研究常常被用于情感分析、数据挖掘、文本分类等,例如微博情感分析、电影评论情感分析、网络商品评论情感分析等。但国内利用情感词汇词典进行情感语言学本体研究还是较少,例如:从计量风格学的角度进行的考察(如刘颖等,2014)^[2]。在对外汉语教学领域,汉语作为第二语言(Chinese as a second language,简称 CSL)学习者的情感词汇研究相对薄弱,有利用语料库进行 CSL 学习者的情感类词偏误研究(郝晓艳,2015)^[3],有对 CSL 学习者褒贬义词习得研究(吴丹,2015)^[4]。还未见到直接利用情感词典对 CSL 学习者进行词汇方面的研究。

因此,本文以情感词汇词典为依托,对中文文本中情感词汇的提取和统计分析,以此为基础采用问卷调查的方式对 CSL 学习者“接受性词汇”^[5]和“产出性词汇”进行考察,旨在研究 CSL 学习者对“接受性词汇”的情感词汇熟悉度测量表现和“产出性词汇”的情感词汇输出表现,并对 CSL 学习者情感词汇认知结果进行描述,提出针对性的建议。

2 中文情感词汇文本反映

2.1 中文情感词汇本体库

中文情感词汇本体库是大连理工大学信息检索研究室在林鸿飞教授带领团队整理和标注的一个中文本体资源。^[6]该资源从不同角度描述一个中文词汇或者短语,包括词语词性种类、情感类别、情感强度及极性等。共含有情感词共计 27466 个^[7]。我们在对比前文提到的各类情感词典,发现该情感词汇知识库标注较为系统、属性信息相对丰富,适合本研究之用。中文情感词汇本体库中,一般的格式如表 1:

表 1 情感词汇本体格式举例

词语	词性种类	词义数	词义序号	情感分类	强度	极性	辅助情感分类	强度	极性
责备	verb	1	1	NN	5	2			
学富五车	idiom	1	1	PH	9	1			
浩气	noun	1	1	PH	5	1			
坦率	adj	1	1	PH	5	1	PB	5	1

在标注中,词性种类一共分为 7 类,分别是名词(noun),动词(verb),形容词(adj),副词(adv),网络词语(nw),成语(idiom),介词短语(pre)。情感强度分为 1, 3, 5, 7, 9 五档,9 表示强度最大,1 为强度最小。极性标注为:0 代表中性,1 代表褒义,2 代表贬义,3 代表兼有褒贬两性。词汇本体中的情感共分为 7 大类 21 小类,如表 2 所示:

表 2 情感词汇分类及示例

编号	情感大类	情感类	例词
1	乐	快乐(PA)	舒服、得意、兴高采烈
2		安心(PE)	知足、安慰、若无其事
3	好	尊敬(PD)	尊严、佩服、肃然起敬
4		赞扬(PH)	高明、准确、高深莫测
5		相信(PG)	稳妥、认可、令人信服
6		喜爱(PB)	喜欢、向往、梦寐以求
7		祝愿(PK)	希望、祝福、万寿无疆
8	怒	愤怒(NA)	脾气、怒视、气急败坏

9	哀	悲伤(NB)	悲哀、可惜、痛哭流涕
10		失望(NJ)	遗憾、叹气、石沉大海
11		疚(NH)	自责、道歉、问心有愧
12		思(PF)	眷恋、关切、魂牵梦萦
13	惧	慌(NI)	紧急、逃亡、迫在眉睫
14		恐惧(NC)	屈服、震惊、令人不安
15		羞(NG)	汗颜、丢人、面红耳赤
16	恶	烦闷(NE)	无聊、难受、心急如焚
17		憎恶(ND)	邪恶、不满、敬而远之
18		贬责(NN)	责备、放荡、无所事事
19		妒忌(NK)	嫉妒、眼馋、嫉贤妒能
20		怀疑(NL)	嫌疑、质问、众说纷纭
21	惊	惊奇(PC)	惊呆、好奇、不可思议

中文情感词汇本体库中，有名词(noun) 3404个，动词(verb) 3460个，形容词(adj) 5261个，副词(adv) 49个，网络词语(nw) 140个，成语(idiom) 14986个，介词短语(preposition) 166个。我们无法从库中直接选取可以用以考察 CSL 学习者认知的词汇，而需要借助中文小说文本提取出汉语母语者所产出(即作者写作)和经常输入(即读者阅读)的情感词汇，以便与 CSL 学习者进行对比。所以，我们需要选择合适的中文文本进行选取，查看中文情感词汇的文本反映。

2.2 四部长篇小说统计分析

我们选取的四部现代汉语长篇小说为：《平凡的世界》、《三体》系列三本、《小时代》系列三本、沈石溪动物小说三本，题材涵盖：当代城乡社会、科幻、军事、现代都市、校园青春、言情、儿童文学等。如表 3：

表 3 四部长篇小说基本信息

简称	书名	作者	出版社	字数	题材
《平凡的世界》	《平凡的世界(第一部)》	路遥	人民文学出版社	近 834800 字	当代城乡社会
	《平凡的世界(第二部)》		人民文学出版社		
	《平凡的世界(第三部)》		人民文学出版社		
《三体》	《三体》	刘慈欣	重庆出版社	近 942000 字	科幻、军事
	《三体II·黑暗森林》		重庆出版社		
	《三体III·死神永生》		重庆出版社		
《小时代》	《小时代 1.0 折纸时代》	郭敬明	长江文艺出版社	近 374000 字	现代都市、校园青春、言情
	《小时代 2.0 虚铜时代》		长江文艺出版社		
	《小时代 3.0 刺金时代》		长江文艺出版社		
沈石溪小说	《一条猎雕的遭遇》	沈石溪	浙江少年儿童出版社	近 390000 字	儿童文学
	《鸟奴》		浙江少年儿童出版社		
	《狼王梦》		浙江少年儿童出版社		

四部小说题材涵盖较广、社会影响较大、销量较高，阅读受众从幼儿到老年人基本能够

涵盖。我们认为所提取出情感词汇的交集，能基本反映现代汉语小说情感词汇用词面貌。

本文对其进行分词、词性标注（使用语料分词和词性标注软件 [CorpusWordParser^{\[1\]}](#)）、统计词频（使用语料字词频率统计工具 [CorpusWordFrequencyApp^{\[2\]}](#)）、删除重复词汇，最后与“情感词汇本体库”取交集抽取情感词汇，从情感分类、词性种类、极性、强度四个角度进行统计，观察其特点以待进一步研究。所抽取的共同情感词汇作为测量材料，对 CSL 学习者认知表现进行对比研究。

四部小说情感词汇的情感分类统计结果如表 4 所示。

表 4 四部小说情感词汇情感分类

情感分类																					
	NA	NB	NC	ND	NE	NG	NH	NI	NJ	NK	NL	NN	PA	PB	PC	PD	PE	PF	PG	PH	PK
《平凡的世界》	52	231	96	172	254	23	20	74	46	6	20	880	315	220	49	140	88	40	85	1063	62
	1.3%	5.8%	2.4%	4.3%	6.4%	0.5%	0.5%	1.8%	1.1%	0.1%	0.5%	22.3%	8.0%	5.5%	1.2%	3.5%	2.2%	1.0%	2.1%	27.0%	1.5%
《三体》	43	155	102	131	150	12	16	47	36	3	21	603	191	133	43	101	70	28	77	818	50
	1.5%	5.4%	3.6%	4.6%	5.2%	0.4%	0.5%	1.6%	1.2%	0.1%	0.7%	21.2%	6.7%	4.6%	1.5%	3.5%	2.4%	0.9%	2.7%	28.8%	1.7%
《小时代》	32	104	61	104	102	15	11	38	24	4	14	431	140	105	23	42	44	15	32	509	31
	1.7%	5.5%	3.2%	5.5%	5.4%	0.7%	0.5%	2.0%	1.2%	0.2%	0.7%	22.9%	7.4%	5.5%	1.2%	2.2%	2.3%	0.7%	1.7%	27.3%	1.6%
沈石溪小说	55	169	104	152	152	16	23	52	40	4	18	646	201	153	43	56	59	24	48	774	38
	1.9%	5.9%	3.6%	5.3%	5.3%	0.5%	0.8%	1.8%	1.4%	0.1%	0.6%	22.8%	7.1%	5.4%	1.5%	1.9%	2.0%	0.8%	1.6%	27.3%	1.3%

就情感词汇的 21 类情感分类来看，四部小说在各类情感词汇的使用比例上较为相似，其中赞扬类（PH）的词汇占比最高，接近 27%，贬责类（NN）次之，接近 23%。这两类情感词就占比接近 50%。快乐类（PA）占比接近 7.5%，悲伤类（NB）占比接近 5.5%，分别排第三和第四。虽然四部小说作者不同、题材不同，但各情感类用词差异较小。

本文对四部小说情感词汇词性种类的统计结果如表 5 所示。

表 5 四部小说情感词汇词性种类

词性种类							
	adj	adv	idiom	noun	nw	prep	verb
《平凡的世界》	1569	20	631	747	7	23	939
	39.8%	0.5%	16.0%	18.9%	0.1%	0.5%	23.8%
《三体》	1216	21	325	518	10	11	729
	42.9%	0.7%	11.4%	18.2%	0.3%	0.3%	25.7%
《小时代》	805	16	301	318	6	8	427
	42.7%	0.8%	16.0%	16.9%	0.3%	0.4%	22.7%
沈石溪小说	1201	21	431	478	10	18	668
	42.4%	0.7%	15.2%	16.9%	0.3%	0.6%	23.6%

四部小说在情感词汇词性种类上较为相似。其中形容词最多，约占 40%，动词次之，约占 23%，名词再次之，约占 17%。三者相对比例基本符合在情感词汇库中的分布。就本次研究而言，形容词将作为考察的重点。

^[1] 肖航.教育部语言文字应用研究所, email: exiaohang@sina.com, website: www.cncorpus.org

^[2] 肖航.教育部语言文字应用研究所, email: exiaohang@sina.com, website: www.cncorpus.org

本文对四本小说情感词汇极性的统计结果如表 6 所示。

表 6 四部小说情感词汇极性

	极性			
	0 (中性)	1 (褒义)	2 (贬义)	3 (兼类)
《平凡的世界》	1272	1471	1179	14
	32.3%	37.3%	29.9%	0.3%
《三体》	922	1062	832	14
	32.5%	37.5%	29.3%	0.4%
《小时代》	618	663	588	12
	32.8%	35.2%	31.2%	0.6%
沈石溪小说	929	992	891	15
	32.8%	35.0%	31.5%	0.5%

在极性方面，四本小说在不同极性的用词比例依旧相似，其中，态度积极的词最多，中性词和态度消极的词数量接近，褒贬兼有的词相对偏少。

表 7 四部小说情感词汇强度

	强度				
	1	3	5	7	9
《平凡的世界》	359	958	1611	723	258
	9.1%	24.3%	40.9%	18.3%	6.5%
《三体》	282	700	1178	485	161
	9.9%	24.7%	41.6%	17.1%	5.6%
《小时代》	182	444	761	339	139
	9.6%	23.6%	40.4%	18.0%	7.3%
沈石溪小说	258	667	1153	519	211
	9.1%	23.5%	40.7%	18.3%	7.4%

在强度方面，5 级强度的词（如：悲伤、粗鲁、端庄、欢乐）最多，约为 40%，1 级（如：安静、迟缓、单一、荒凉）和 9 级（如：暴怒、荒唐、完美、心花怒放）强度的词最少，约为 9% 和 6%。四部小说情感词汇的强度分布基本上符合正态分布图。

总体来看，我们发现虽然四部小说题材不同、作者不同，但在用词上非常相近。即使字数相差悬殊的小说，在情感分类、词性种类、极性和强度的各项用词比例仍然相近。由此可以推断出，现代汉语长篇小说在情感词汇的使用上，并不会因为作者不同而产生较大差异。这一结论可以为计量风格学的文本分析提供参考。

四部小说的情感词汇用词占比非常接近，是否因为受制于情感词汇本体库的因素限制呢？由此本文统计了四部小说的总字数、总用词量和词性占比，发现结果同情感词汇统计的结果相似，四部小说总用词比例相近，不存在外部因素影响。

2.3 四本小说共同情感词汇提取

为了深入了解四部小说情感词语的情况，以便考察 CSL 学习者的认知，本文对四本小说共同情感词汇进行了提取，得到了 784 个词，在考察 CSL 学习者接受性词汇时，考虑到测量效果和情感词汇的特点，我们只选取形容词作为测量对象。共有 409 个形容词。在考察 CSL 学习者产出性词汇时，我们将详细对比。具体情况见下文。

3 CSL 学习者接受性情感词汇考察

上文中四本书提取出情感词汇是汉语母语者所产出（作者写作）和经常输入（即读者阅读）的词，是汉语母语者非常熟悉的词。但对于 CSL 学习者来说是否熟悉有待于进一步考察。

针对这一问题，我们将测量材料（即从四部小说提取的情感词汇）制作成测试词表，对 CSL 学习者进行熟悉度测试，并对熟悉度结果进行检验。

首先是词表制作。我们将上文提到的四部小说所抽取的 409 个形容词，按照频率排序，等比提取 30 个词（等比间距为 13，例如，取第 1 个，第 14 个，第 27 个……以此类推），如表 8：

表 8 情感词汇形容词按比例抽取 30 词词表

序号	词语	序号	词语	序号	词语
1	迅速	11	惊慌	21	自豪
2	奇怪	12	可笑	22	优美
3	漫长	13	高贵	23	甜美
4	干净	14	寂寞	24	暗淡
5	亲切	15	急促	25	闷热
6	必要	16	嘈杂	26	贸然
7	高大	17	丑陋	27	火热
8	肮脏	18	自卑	28	悲凉
9	惊恐	19	天真	29	厌倦
10	焦急	20	无聊	30	纯洁

针对这 30 个词，我们设计了六级量表，熟悉度从 1 到 6，1 表示最熟悉，6 表示最不熟悉。表头信息为：国籍、性别、年级、HSK 等级。其中，年级和 HSK 等级共同作为被试汉语水平的评判标准。我们在北京语言大学范围内进行问卷收集，共收集到来自 14 个国家的有效问卷 30 份，其中高级 5 份，中级 20 份，低级 5 份。CSL 中级水平学习者是本次实验重点考察的对象，因为他们能够基本反映 CSL 学习者对汉语词汇的认知特点，能够反映从初级向中级过渡、中级向高级过渡时词汇学习的问题，符合本次实验预期考察的目的。

接下来，我们进行熟悉度测量，并对测量结果进行检验。

科勒-拉普假设认为，人们对某些词的熟悉程度和其出现的频率之间存在某种关系，可以假设：人们越熟悉一个词，其出现频率就越高。那是否对于 CSL 学习者也是如此？

所以，我们针对 CSL 学习者检验科勒-拉普假设，即：

$$y = \frac{V}{1+Ax^B},$$

其中，V 表示数据中最大熟悉度的值，x 表示词的频率，A、B 是函数的参数。^[8]

考虑到初级 CSL 学习者和高级 CSL 学习者的水平特点，不适合列入熟悉度计算，所以我们只对其进行描述，不进行计算。

我们用提取出的这 30 个词语在四本小说中的频率为 x，对 20 名中级 CSL 学习者问卷熟悉值进行统计，计算出熟悉度均值为 y。y 值为 1 是最熟悉，y 值为 6 是最不熟悉。V 值为 6。结果如下表 9 所示。

表 9 测量词表频率与 20 名中级 CSL 学习者熟悉度

序号	词语	频率序号	x	y	序号	词语	频率序号	x	y
1	奇怪	2	0.0099	2	16	火热	27	0.0007	4.05
2	干净	4	0.0063	2.1	17	急促	15	0.0021	4.1
3	高大	7	0.0043	2.2	18	惊恐	9	0.0035	4.25
4	必要	6	0.0048	2.25	19	肮脏	8	0.0039	4.3
5	高贵	13	0.0023	2.35	20	焦急	10	0.0032	4.35
6	可笑	12	0.0026	2.5	21	自卑	18	0.0017	4.4
7	无聊	20	0.0014	2.55	22	悲凉	28	0.0006	4.45
8	寂寞	14	0.0022	2.75	23	惊慌	11	0.0028	4.5
9	亲切	5	0.0054	2.85	24	暗淡	24	0.001	4.55
10	漫长	3	0.0073	2.95	25	闷热	25	0.001	4.6
11	优美	22	0.0013	3.05	26	贸然	26	0.0008	4.9
12	甜美	23	0.0011	3.05	27	嘈杂	16	0.0019	5.05
13	自豪	21	0.0014	3.35	28	纯洁	30	0.0005	5.05
14	迅速	1	0.0193	3.65	29	丑陋	17	0.0017	5.15
15	天真	19	0.0015	3.85	30	厌倦	29	0.0006	5.25

观察结果可以看到，按熟悉度降序排列后，熟悉度最高的前 15 个词，有 10 个词频率降序排列在前 15，5 个词频率降序排列在后 15。

为了更精确地得出结论，我们继续进行科勒-拉普假设检验。使用 MATLAB 中 inline 自定义函数模型，与数据拟合后求得参数 A, B (其中 V=6)：

(1) 使用 inline 函数定义模型，三个参数分别为：函数模型，待定系数，自变量。所以，`myfunc = inline('6./(1+beta(1).*x.^beta(2))', 'beta', 'x');`；

(2) 待定系数的预估值：`beta0 = [0.1, 0.1]'`；

(3) 数据集和函数模型进行拟合：`beta = nlinfit(x, y, myfunc, beta0)`。

其中 `beta(1)=10.3591`, `beta(2)=0.4566`, 则 `A=10.3591`, `B=0.4566`。

将所得结果可视化，显示数据散点图及拟合曲线，计算函数模型的拟合优度 R^2 ，对比分析得出结果， $R^2=0.2946$ ：

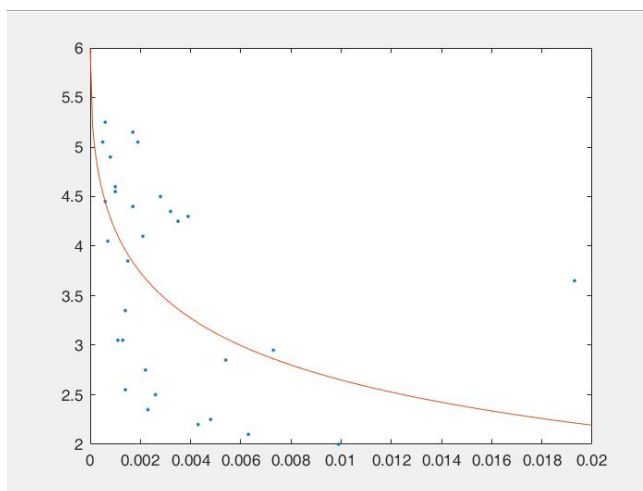


表 10 数据拟合图

结果显示, 拟合优度 $R^2=0.2946$, 结果较差, 不能证明科勒-拉普假设针对 CSL 学习者在情感词汇方面的假设, 即 CSL 学习者越熟悉一个情感词汇, 其出现的频率不一定越高。也就是说母语学习者最熟悉的词, 不一定 CSL 学习者也熟悉。

究其原因, CSL 学习者的词汇大部分来源于教材和教师的词汇教学, 其情感词汇的学习和使用也受制于教材和教师。CSL 学习者对频率高的情感词汇熟悉度不够, 将会使 CSL 学习者在阅读现代汉语小说时具有更大的困难。在对外汉语教材的编写上, 尤其是阅读课, 应该关注现代汉语小说中高频的情感词汇。这对提升 CSL 学习者的在阅读和对话时的情感判断有重要帮助, 进而使其在语义理解上更快更准确。

4 CSL 学习者产出词汇统计分析

前文利用情感词典和文本对 CSL 学习者接受性情感词汇进行了测试性考察, 接下来我们从 CSL 学习者产出的角度进行情感词汇输出考察。一般而言, 接受性词汇和产出性词汇之间存在巨大差距, 也就是所谓“产出性词汇门槛”, CSL 学习者的产出性词汇量远远小于接受性词汇量。⁹⁾但是否在情感词汇上也是如此? CSL 学习者在产出词汇以及产出情感词汇上有什么特点? 这些特点对对外汉语教学有什么启发? 我们设计了问卷进行调查研究。

考虑到 CSL 学习者产出词汇的难度和我们预期产出词汇的广度, 本文不将 CSL 初级和准中级学习者列入考察范围。我们认为对 CSL 中高级学习者的考察能够反映出 CSL 学习者产出词汇时特点和问题的全貌。

CSL 学习者产出词汇表问卷调查对象为北京语言大学留学生, HSK 等级 4 级以上, 学习汉语时间 4 年到 6 年, 整体水平为中高级。共收集到有效问卷 32 份, 来自韩国、日本、塞尔维亚、葡萄牙、埃及等多个国家。产出词汇表采取闭卷方式, CSL 学习者自由输出 300 个词。我们将其校对、电子化、进行描述, 并抽取其中的情感词汇进行统计分析。

据统计, 除去无法识别的错别字和生造词, 共收集到有效词汇 9674 个, 删除重复项后为 3549 个。对 CSL 学习者产出情感词汇进行抽取, 共取得词语 372 个。

考虑到情感词汇的特点, 我们同样重点观察形容词的产出情况。

(1) 按情感分类

表 11 CSL 学习者产出情感词汇情感分类计数

情感分类	NA 愤怒	NB 悲伤	NC 恐惧	ND 憎恶	NE 烦闷	NG 羞耻	NH 疚	NI 慌	NJ 失望	NL 怀疑	NN 贬责	PA 快乐	PB 喜爱	PC 惊奇	PD 尊敬	PE 安心	PF 思	PG 相信	PH 赞扬	PK 祝愿
CSL 学习者产出情感词汇数量	3	12	9	10	19	7	2	5	4	1	54	44	30	3	19	13	6	19	104	8
CSL 学习者产出情感词汇形容词数量	2	4	4	2	13	6	2	2	2	0	24	19	7	1	3	6	0	6	63	1
四本书同情感词汇数量	8	49	24	37	42	5	7	15	10	3	131	76	46	17	15	21	9	23	230	16

根据表 11 的数据, 我们发现, 双方都没有嫉妒 (NK) 这一类, 这类词在情感词汇本体库中也只有 43 个, 例如“眼馋”、“宠儿”、“醋意”等。

根据表 4 对四本小说的统计, 我们发现赞扬类 (PH) 的词汇占比最高, 接近 27%, 贬责类 (NN) 次之, 接近 23%。这两类情感词就占比接近 50%。快乐类 (PA) 占比接近 7.5%, 悲伤类 (NB) 占比接近 5.5%, 分别排第三和第四。但在 CSL 学习者产出的情感词汇中, 我们可以看到, 不同情感分类按照数量排序, 前九种为: 赞扬 (PH)、贬责 (NN)、快乐 (PA)、喜爱 (PB)、尊敬 (PD)、相信 (PG)、烦闷 (NE)、安心 (PE)、悲伤 (NB)。两者相比大体相近, 不同在于 CSL 学习者对悲伤类的词语产出较少。在实验结果反馈和分析时, 我们认为其受外部因素影响, 如汉语教学中教师和教材积极乐观的态度、填写问卷时轻松愉快

的心情等，所以 CSL 学习者在产出词汇过程中对悲伤（NB）类词汇输出较少，可能并不是因为相关词汇量小的原因。针对这类词（悲伤类）的二语教学是否应该以母语者的掌握水平为标准？这一点值得继续深入研究。

（2）极性

表 12 CSL 学习者产出情感词汇极性计数

极性	极性	CSL 学习者产出 词汇数量	CSL 学习者产出 形容词数量	四本书共同情感 词汇	CSL 学习者产出 词汇占比（近似 值）	CSL 学习者产出 形容词占比（近 似值）	四本书共同情感 词汇占比（近似 值）
0	中性	150	66	313	40.32%	39%	40.13%
1	褒义	149	75	297	40.05%	44.30%	38.08%
2	贬义	73	28	170	19.62%	16.50%	21.79%
3	褒贬兼有	0	0	4	0	0	0.51%

表 12 中，CSL 学习者情感词汇极性占比相似，差异不大。四本书共同情感词汇中，褒贬兼有的词为“好事”、“解除”、“迷恋”、“天真”。这类词由于数量较少，对情感判断影响较大。在实际对外汉语教学中，我们建议重点强调其不同用法。

（3）强度

表 13 CSL 学习者产出情感词汇强度计数

强度	强度备注	CSL 学习者产出 情感词汇数量	CSL 学习者产出 情感词汇形容 词数量	四本书共同情感 词汇数量	CSL 学习者产出 词汇占比（近似 值）	CSL 学习者产出 词汇形容词占 比（近似值）	四本书共同情 感词汇占比（近 似值）
1	最小	44	24	79	11.89%	14.20%	10.15%
3		97	49	220	26.22%	28.90%	28.28%
5		170	69	315	45.95%	40.80%	40.49%
7		49	20	119	13.24%	11.80%	15.3%
9	最大	10	6	45	2.7%	3.5%	5.78%

在强度方面，CSL 学习者产出的强度和四本书共同情感词汇相似。在 5 级强度上，CSL 学习者占比高于四本书共同情感词汇，但 9 级强度上，CSL 学习者占比低于四部小说共同情感词汇。这说明 CSL 学习者对表达强烈感情的词汇掌握的还不够多。

5 结语

通过利用中文情感词汇本体库对文学文本和 CSL 学习者产出性词汇的对比分析，我们发现，现代汉语长篇小说在情感词汇的使用上，并不会因为作者、题材、内容不同而产生较大差异。这一结论可以为计量风格学的文本分析提供参考。就文本中情感词汇的情感分类来看，赞扬类和贬责类两类情感词占总词数的一半，其它 19 类情感词占一半，而且积极正向的词数量多于消极负向的词。

另外，我们发现 CSL 学习者对频率高的情感词汇熟悉度不够，解决这个问题对他们在文本阅读和对话时的情感判断有很大帮助。中高级 CSL 学习者在产出性词汇中，CSL 学习者对表达强烈感情的词汇掌握的不够多。

因此我们建议，在对外汉语教材的编写上，尤其是阅读课，应该关注现代汉语小说中高频的情感词汇。这对提升 CSL 学习者的在阅读和对话时的情感判断有一定帮助，进而使其在语义理解上更快更准确。在具体的教学中，我们可以强化他们在阅读现代汉语小说时对词的情感理解的训练，从而更加熟悉现代汉语中的词语使用特点和中国人的情感表达方式，以便突破汉语水平由“高级”到“地道”的瓶颈期。同时，我们也应该注意加强 CSL 学习者褒贬兼有词语的学习，在表达情感方面，加强对表达强烈感情词语的使用。

参考文献

- [1] 李军. 中文评论的褒贬义分类试验研究[D]. 北京:清华大学, 2008.
- [2] 刘颖, 肖天久. 《红楼梦》计量风格学研究[J]. 红楼梦学刊, 2014 (4): 260-281
- [3] 郝晓艳. 基于 HSK 动态作文语料库的情感类词偏误研究[D]. 山西大学 2015
- [4] 吴丹. 留学生汉语褒贬义词习得研究[D]. 上海师范大学 2015
- [5] 莫丹. 美国学习者汉语写作产出性词汇知识的发展路径考察[J]. 海外华文教育, 2017 (5): 579-605
- [6] 徐琳宏, 林鸿飞, 潘宇, 任惠, 陈建美. 情感词汇本体的构造[J]. 情报学报, 2008 (2): 180-185
- [7] 郑毅. 基于情感词典的中文微博情感分析研究[D]. 中山: 中山大学, 2014
- [8] 刘海涛. 计量语言学导论[M]. 北京: 商务印书馆, 2017
- [9] 莫丹. 华裔与非华裔汉语学习者产出性词汇知识差异及其对写作质量的影响[J]. 云南师范大学学报 (对外汉语教学与研究版), 2015 (5): 33-41

作者联系方式: 张易扬; 北京市海淀区学院路北京语言大学; 100083; 18811167732;
bournezyy@163.com