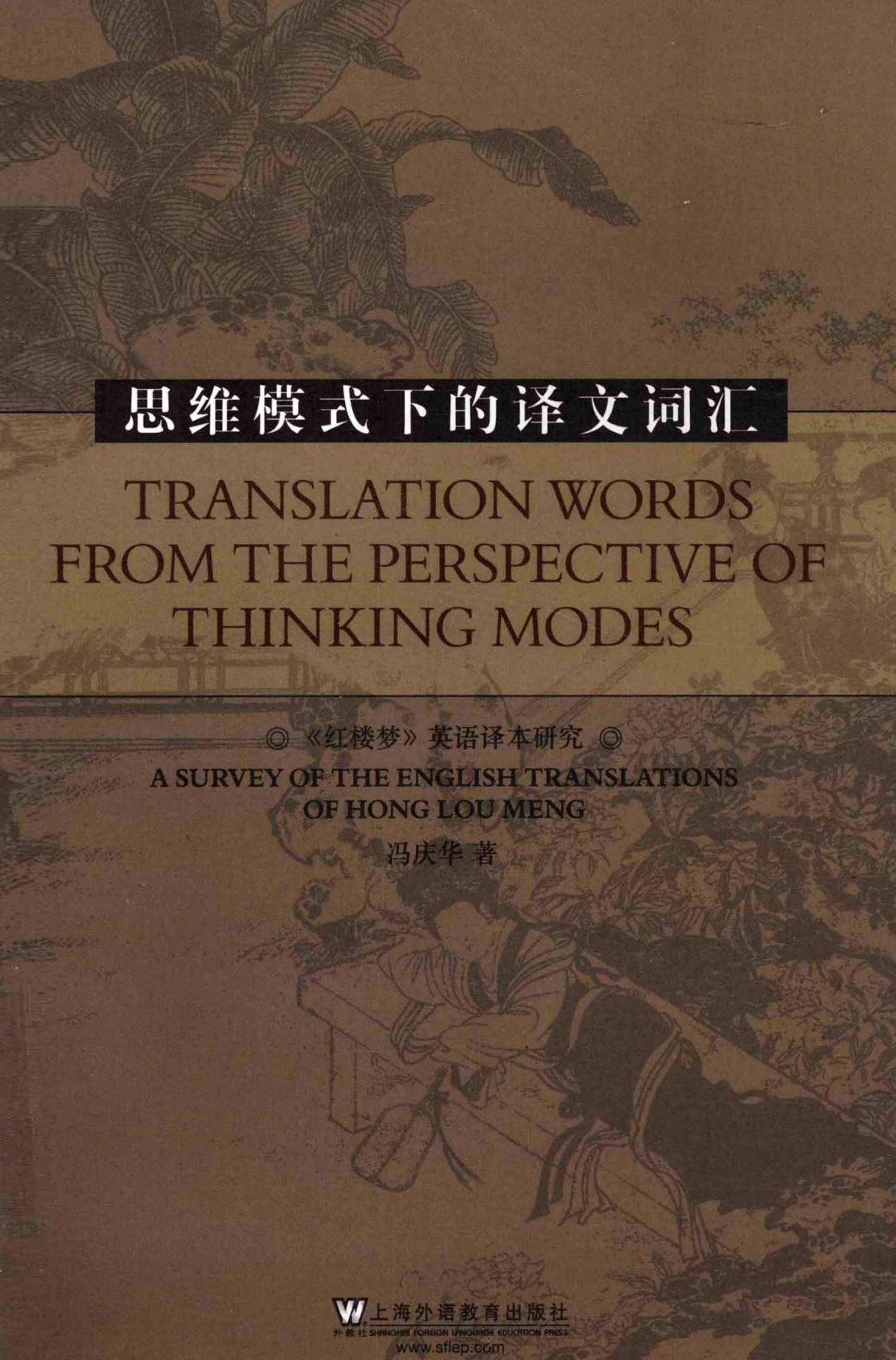




思维模式下的译文词汇

TRANSLATION WORDS FROM THE PERSPECTIVE OF THINKING MODES

◎ 《红楼梦》英语译本研究 ◎

A SURVEY OF THE ENGLISH TRANSLATIONS
OF HONG LOU MENG

冯庆华 著

思维模式对译文风格及译文词汇有着很大的影响。本作品重点研究的是东西方思维模式对译文词汇及其搭配的作用。我们通过文本分析软件对西方译者与中国译者翻译同一部中国文学作品的英译文本进行分析,发现他们在词汇使用及其搭配上存在着明显的差异,那些高频词、特色词、独特词,那些以抽象名词充当的主语与宾语,那些独特的形容词与副词,都充分显示了西方思维模式对英语译文的作用和影响,从而形成一定的语言特点。中国思维下的译者应该从这些西方思维模式下的语言特点得到一些启示,使我们自己英语译文的语言更加地道,增强译文的可读性,更好地向世界传播和推广我们的中国文化。

ISBN 978-7-5446-2683-5



9 787544 626835 >

定价: 78.00 元

思维模式下的译文词汇

TRANSLATION WORDS
FROM THE PERSPECTIVE OF
THINKING MODES

◎ 《红楼梦》英语译本研究 ◎

A SURVEY OF THE ENGLISH TRANSLATIONS
OF HONG LOU MENG

图书在版编目(CIP)数据

思维模式下的译文词汇/ 冯庆华著. —上海: 上海外语教育出版社, 2012
ISBN 978-7-5446-2683-5

I. ①思… II. ①冯… III. ①英语-翻译-研究 IV. ①H315.9

中国版本图书馆 CIP 数据核字(2012)第 021054 号

出版发行: **上海外语教育出版社**

(上海外国语大学内) 邮编: 200083

电 话: 021-65425300 (总机)

电子邮箱: bookinfo@sflep.com.cn

网 址: <http://www.sflep.com.cn> <http://www.sflep.com>

责任编辑: 苗 杨

印 刷: 上海信老印刷厂

开 本: 787×965 1/16 印张 31.75 字数 565 千字

版 次: 2012 年 5 月第 1 版 2012 年 5 月第 1 次印刷

印 数: 1 100 册

书 号: ISBN 978-7-5446-2683-5 / H · 1279

定 价: 78.00 元

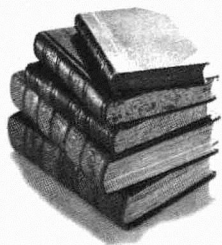
本版图书如有印装质量问题, 可向本社调换



目 录

0 绪论	1
1 高频词	5
1.1 高频名词	24
1.2 高频形容词	43
1.3 高频动词	46
1.4 高频副词	71
1.5 高频介词	78
1.6 高频口语词	96
1.7 高频书面词	107
1.8 高频指示代词	108
2 特色词	112
2.1 特色名词	121
2.2 特色形容词	144
2.3 特色动词	159
2.4 特色副词	169
2.5 特色介词	177
2.6 特色口语词	183
2.7 特色书面词	204
3 独特词	224
3.1 独特名词	256
3.2 独特形容词	289
3.3 独特动词	300
3.4 独特副词	312

3.5	独特介词	317
3.6	独特书面词	328
4	词语搭配	337
4.1	名词的修饰词语和谓语	352
4.2	动词的修饰词语和主语、宾语	371
4.3	形容词的被修饰对象	375
4.4	副词的搭配	378
5	否定词	383
5.1	原著否定词研究	383
5.2	译文否定词研究	405
5.3	译文否定词搭配分析	439
	结语	451
	附录一 《三国演义》人民文学出版社版字频逆序表	454
	附录二 《水浒传》人民文学出版社版字频逆序表	462
	附录三 《西游记》人民文学出版社版字频逆序表	471
	附录四 《红楼梦》人民文学出版社版字频逆序表	480
	附录五 《毛泽东选集》人民出版社版字频逆序表	488
	附录六 《红楼梦》与《毛泽东选集》比较的独特汉字	496
	参考书目	500



0

绪 论

对《红楼梦》原著与译文的词汇进行比较研究可以帮助我们挖掘并发现各种思维模式对语言风格的影响,从而可以总结这些风格产生的主要因素,给我们学习或树立一种写作或翻译风格提供一些重要依据。

我们选择《红楼梦》原著及其两个英语全译本作为我们的主要研究对象,因为该原著是公认优秀的文学作品,两个英语全译本是公认优秀的翻译作品。

在选择《红楼梦》英语译本时,我们毫不犹豫地选择了大卫·霍克斯与约翰·闵福德的译本以及杨宪益与戴乃迭的译本,因为这两个译本有共性:最大量化的忠实,最大量化的通顺。同时,这两个优秀译本又有各自的个性:霍译倾向于交际翻译法,杨译倾向于语义翻译法。

交际翻译法与语义翻译法都是翻译实践与研究两个极为重要的方法。交际翻译法更注重灵活性,但没有极端自由编译的特性;语义翻译法更注重与原文语言文化形式上的一致,但绝对没有逐字逐行翻译的拘谨。

交际翻译强调的是与译文在效果上的等同;而语义翻译强调的是与原文在内容上的等同。

交际翻译的译文通常是通顺易懂,行文流畅;语义翻译的译文通常是规范自然,生动形象。

为了追求译文的效果,霍译经常会为读者对原文语言中的文化进行因地制宜的改装和归化;为了尽力保持原文的风

貌,杨译经常会为读者提供原汁原味的中国文化形式与修辞习惯,也许后者对其中不少西方读者来说还不太熟悉。

霍译的译者竭尽全力,为自己的母语读者再现《红楼梦》原作所有的优美,不想失去一点一滴。

霍克斯先生在其《红楼梦》英译本前言中发表了这样一段关于处理“红”字的感言:

One bit of imagery which Stone-enthusiasts will miss in my translation is the pervading *redness* of the Chinese novel. One of its Chinese titles is red, to begin with, and red as a symbol — sometimes of spring, sometimes of youth, sometimes of good fortune or prosperity — recurs again and again throughout it. Unfortunately — apart from the rosy cheeks and vermeil lip of youth — redness has no such connotations in English and I have found that the Chinese reds have tended to turn into English golds or greens (‘spring the green spring’ and ‘golden girls and boys’ and so forth). I am aware that there is some sort of loss here but have lacked the ingenuity to avert it.

霍克斯先生深感中国文化广博深奥,要翻译所有的文化因子必有难度,他在《红楼梦》英译本前言谈了他的翻译原则:

In translating this novel I have felt unable to stick faithfully to any single text. I have mainly followed Gao E's version of the first chapter as being more consistent, though less interesting, than the other ones; but I have frequently followed a manuscript reading in subsequent chapters, and in a few, rare instances I have made small emendations of my own. **My one abiding principle has been to translate everything — even puns.** For although this is, in the sense I have already indicated, an ‘unfinished’ novel, it was written (and rewritten) by a great artist with his very lifeblood. I have therefore assumed that whatever I find in it is there for a purpose and must be dealt with somehow or other. I cannot pretend always to have done so successfully, but if I can convey to the reader even a fraction of the pleasure this Chinese novel has given me, I shall not have lived in vain.

本作力图从词法角度去比较研究优秀汉语原作与优秀英语译作在语言结构上的共性与差异,去挖掘分析不同思维模式下的译者风格。

本作在统计分析中将借助几个文本分析软件,这几个软件分别是:

Concordance 3.0(简称 Concordance); Concapp Windows Application 4.0(简称 Concapp); Antconc 3.2.1w; Antwordprofiler 1.200w。下面是这四个软件的相关信息:

Product Name	Concordance	Concapp Windows Application	Antconc	Antwordprofiler
Product Version	3.2	4.0	3.2.4w (Windows)	1.200w (Windows)
Company Name	RJCW		Waseda University Japan	Waseda University Japan
Legal Copyright	R. J. C. WATT 1999, 2000, 2002.	Chris Greaves 1993-2003.	Laurence Anthony 2011.	Laurence Anthony 2009.

除此之外,我们还将通过北京大学中文系研发的“现代汉语文本预处理”(TextPreProcessing 1.0, 2001)和“词语分隔”(WordSeg 1.0, 2002)来处理汉语文本的词语分隔问题。这两个软件可在北京大学汉语语言学研究 中心(<http://ccl.pku.edu.cn/>)网上下载。经过词语分隔处理的汉语文本就可以像英语文本一样进行文本统计分析了。

当然,文中许多资料筛选与统计也大量依靠了其他文字与报表处理软件。

为了对原作和译作进行横向的文本比较分析,我们使用了大量的汉语文本与英语文本。

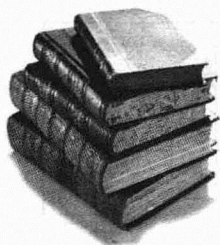
汉语文本包括:《三国演义》、《水浒传》、《西游记》、《金瓶梅》、《红楼梦》、《鲁迅选集》、《家》、《子夜》、《围城》、《长恨歌》、《毛泽东选集》。

英语文本包括:“霍译红楼梦”、“杨译红楼梦”、“英国小说一”、“英国小说二”、“美国小说”、“纪实文学”、“毛选英文版”。

下面是我们引用的英语作品的基本统计信息:

文件名	总词数	词汇量	平均词频
霍译红楼梦	831 414	22 557	36.858
杨译红楼梦	626 123	16 864	37.128
英国小说一	807 173	22 088	36.544
英国小说二	873 521	26 208	33.331
美国小说	844 286	24 560	34.376
纪实文学	1 203 195	36 748	32.742
毛选英文版	672 884	14 433	46.621

“霍译红楼梦”是指霍克斯与闵福德的《红楼梦》英译本；“杨译红楼梦”是指杨宪益与戴乃迭的《红楼梦》英译本；“英国小说一”是指《大卫·科波菲尔》、《傲慢与偏见》、《简·爱》、《呼啸山庄》和《苹果树》等五部英国小说原著合集；“英国小说二”是指《名利场》、《德伯家的苔丝》、《雾都孤儿》、《鲁宾逊漂流记》和《双城记》等五部英国小说原著合集；“美国小说”是指《飘》、《嘉莉妹妹》、《汤姆·索亚历险记》、《哈克贝利·费恩历险记》和《红字》等五部美国小说原著合集；“纪实文学”是指《第三帝国兴亡》和《光荣与梦想》等两部英语纪实文学原著合集；“毛选英文版”是指《毛泽东选集》第一卷至第五卷由中国北京外文出版社出版的英译本合集，英译委员会主任委员钱钟书负责英文翻译、审稿、定稿工作。



所谓高频词就是在一个文本里使用频率特别高的词语。用一个词语在整个文本中所占的百分比或者该词语在整个词频中所处的前后位置来决定该词语是否为高频词。每个文本由于大小不一样,因此选择或定义相关文本的高频词的百分比或排序位置也会有所区别。

各时期小说字频词频比较

对照数百年来不同时期写下的《红楼梦》等名著,现当代的文学作品在语言上发生了很大的变化,我们通过字频统计比较可以明显感到这一点。下面,我们选择了600多年前的《水浒传》(施耐庵著)、400多年前的《西游记》(吴承恩著)、200多年前的《红楼梦》(曹雪芹著、高鹗续)、80年前的《子夜》(茅盾著)、17年前的《长恨歌》(王安忆著)等五个文本,把不同时代、不同文体的前20个常用字做成表格,供我们比较研究之用。表中所列次数为相关字在该文本中使用的次数,字频为该字的使用或出现频率,也就是其使用次数占该文本总字数的百分比。主要人物的人名用字不计在内,高亮部分为该部作品在五部作品排名前20位汉字中的独特字。

	水浒传 600 多年前			西游记 400 多年前			红楼梦 200 多年前			子夜 80 年前			长恨歌 17 年前		
	字	次数	字频	字	次数	字频	字	次数	字频	字	次数	字频	字	次数	字频
1	了	11 459	1.359	道	11 080	1.549	了	21 368	2.442	的	8 635	3.313	的	11 316	4.237
2	道	10 433	1.237	不	8 882	1.242	不	15 129	1.729	了	5 446	2.089	是	7 239	2.711
3	一	10 029	1.189	一	7 968	1.114	的	14 975	1.711	一	4 914	1.885	一	5 301	1.985
4	来	9 798	1.162	了	7 749	1.083	一	12 230	1.398	是	4 139	1.588	了	4 601	1.723
5	人	8 828	1.047	那	7 516	1.051	来	11 534	1.318	不	3 556	1.364	不	4 229	1.584
6	不	8 351	0.990	我	7 224	1.010	道	11 030	1.261	他	2 957	1.135	有	3 415	1.279
7	个	6 577	0.780	是	6 497	0.908	人	10 591	1.210	来	2 409	0.924	她	2 799	1.048
8	是	5 942	0.704	来	5 994	0.838	是	10 225	1.169	在	2 317	0.889	这	2 600	0.974
9	上	5 754	0.682	他	5 751	0.804	说	9 771	1.117	就	2 093	0.803	人	2 323	0.870
10	去	5 496	0.652	个	5 704	0.797	我	9 266	1.059	那	2 077	0.797	他	2 301	0.862
11	里	5 404	0.641	你	5 522	0.772	这	7 858	0.898	有	2 039	0.782	也	2 244	0.840
12	得	5 243	0.622	的	5 425	0.758	他	7 718	0.882	这	2 000	0.767	上	2 104	0.788
13	那	4 849	0.575	有	4 487	0.627	你	7 173	0.820	我	1 897	0.728	来	2 058	0.771
14	下	4 815	0.571	大	4 231	0.592	去	6 217	0.711	着	1 762	0.676	在	1 976	0.740
15	大	4 644	0.551	得	3 845	0.538	也	6 171	0.705	人	1 731	0.664	说	1 955	0.732
16	在	4 581	0.543	这	3 807	0.532	着	6 133	0.701	们	1 640	0.629	里	1 937	0.725
17	我	4 513	0.535	上	3 725	0.521	有	6 037	0.690	上	1 632	0.626	着	1 831	0.686
18	军	4 487	0.532	去	3 699	0.517	儿	6 030	0.689	里	1 553	0.596	就	1 659	0.622
19	只	4 362	0.517	见	3 429	0.479	个	5 694	0.651	地	1 552	0.595	个	1 651	0.619
20	见	4 276	0.507	老	3 412	0.477	子	5 484	0.627	说	1 526	0.584	都	1 402	0.525

在上面这个表格中,我们发现五个字是在五个文本里全都出现了,一个是“了”字,排位都在前四位,一个是“一”字,排位也都在前四位,一个是“不”字,排位都在前六位,一个是“是”字,排位在前八位,而且逐渐上升,还有一个是“来”字,排位先升后降。

我们注意到评判文本性质的关键字“的”在不同时期与不同文体中有不同的字频和排位。《水浒传》里的“的”字出现了 4 112 次,字频为 0.488%,排位第 24;《西游记》里的“的”字出现了 5 425 次,字频为 0.758%,排位第 12;《红楼梦》里的“的”字出现了 14 975 次,字频为 1.711%,排位第三;《子夜》里的“的”字出现了 8 635 次,字频为 3.313%,排位第一;《长恨歌》里的“的”字出现了 11 316 次,字频为 4.237%,排位第一。“的”字发展非常明显,尤其是五四运动以后,白话文取代文言文,汉语的语风出现了极大的变化。“的”字的发展实际上是汉语发展的一个缩影。再需要说明的一点是,《长恨歌》排前两位的是明显口语常用字“的”与“是”字,这与其女性作者和女性题材不无关系。

《红楼梦》与《毛泽东选集》字频词频比较

曹雪芹著作、高鹗续写的120回本《红楼梦》(蔡义江校本)共有736 023个汉字,总字类4 275,平均单字使用频率为172次。

《毛泽东选集》第一卷至第五卷共有895 881个汉字,总字类3 295,平均单字使用频率为272次。

如果我们以纯粹的字为单位做字频统计,《红楼梦》文本列前100位的汉字为:

了;不;的;一;来;道;人;是;说;我;这;他;你;去;也;着;玉;有;儿;宝;个;子;又;贾;里;们;见;只;得;那;太;便;好;在;笑;家;上;么;大;姐;头;听;就;出;回;知;要;日;下;都;心;二;事;老;过;还;话;起;自;如;看;没;叫;两;到;母;些;时;之;今;小;问;因;奶;等;凤;娘;可;什;呢;忙;想;夫;面;爷;才;中;王;打;此;进;罢;倒;样;吃;和;姑;正;后;无。

《毛泽东选集》第一卷至第五卷列前100位的汉字为:

的;是;一;不;国;人;在;民;们;有;这;和;中;主;了;我;个;地;要;战;党;大;义;就;以;产;为;作;对;会;争;政;他;军;上;于;方;动;阶;来;能;反;分;工;年;革;命;级;农;之;也;而;日;同;经;全;到;种;发;可;本;些;部;十;力;社;多;生;共;现;事;面;三;所;还;得;过;说;行;进;实;自;问;成;内;敌;但;没;资;其;去;题;后;都;二;利;解;那;必。

对照这两个字频表,我们发现两者重合的只有42个:

的;是;一;不;人;在;们;有;这;和;中;了;我;个;要;大;就;他;上;来;时;之;也;日;到;可;些;事;面;还;得;过;说;进;自;问;没;去;后;都;二;那。

两者不一样的竟然有58个。对照研究下面两个特色汉字表对我们研究两个文本是很有启发的。

《红楼梦》的58个特色汉字为:

道;你;着;玉;儿;宝;子;又;贾;里;见;只;太;便;好;笑;家;么;姐;头;听;出;回;知;下;心;老;话;起;如;看;叫;两;母;今;小;因;奶;等;凤;娘;什;呢;忙;想;夫;爷;才;王;打;此;罢;倒;样;吃;姑;正;无。

《毛泽东选集》的58个特色汉字为:

国;民;主;地;战;党;义;以;产;为;作;对;会;争;政;军;于;方;动;阶;能;反;分;工;年;革;命;级;农;而;同;经;全;种;发;本;部;十;力;社;多;生;共;现;三;所;行;实;成;内;敌;但;资;其;题;利;解;必。

我们也可以对《红楼梦》文本进行词语分隔,然后对其进行词频统计。我们根据不同词长(phrase length)的词汇分别进行排列,列出各自的前

20 位。

《红楼梦》的一字词汇共计 3 881 种(types), 占总词汇种类 20 004 的 19.4%; 出现次数总计 377 588 次(occurrences), 占总词汇使用次数 546 327 的 69.1%。词频排列前 20 个一字词汇分别为:(1)了;(2)的;(3)道;(4)我;(5)是;(6)他;(7)说;(8)不;(9)你;(10)一;(11)也;(12)来;(13)着;(14)又;(15)去;(16)便;(17)人;(18)这;(19)有;(20)笑。

《毛泽东选集》一至五卷的一字词汇共计 2 520 种(types), 占总词汇种类 15 755 的 16%; 出现次数总计 257 192 次(occurrences), 占总词汇使用次数 552 409 的 46.6%。词频排列前 20 个一字词汇分别为:(1)的;(2)是;(3)和;(4)了;(5)在;(6)不;(7)一;(8)有;(9)这;(10)要;(11)也;(12)就;(13)地;(14)人;(15)中;(16)上;(17)十;(18)都;(19)年;(20)必。

两者重合的一字词汇只有 8 个:的;了;人;是;也;一;有;这。去掉这 8 个一字词汇, 其余 12 个都是各自文本的特色词汇。

《红楼梦》的二字词汇共计 10 593 种(types), 占总词汇种类 20 004 的 52.9%; 出现次数总计 154 057 次(occurrences), 占总词汇使用次数 546 327 的 28.2%。词频排列前 20 个二字词汇分别为(不含人名与称谓):(1)什么;(2)一个;(3)我们;(4)不知;(5)你们;(6)如今;(7)出来;(8)起来;(9)说道;(10)这里;(11)众人;(12)那里;(13)自己;(14)一面;(15)只见;(16)怎么;(17)没有;(18)这个;(19)知道;(20)听见。

《毛泽东选集》一至五卷的二字词汇共计 10 099 种(types), 占总词汇种类 15 755 的 64.1%; 出现次数总计 261 148 次(occurrences), 占总词汇使用次数 552 409 的 47.3%。词频排列前 20 个二字词汇分别为:(1)我们;(2)中国;(3)他们;(4)人民;(5)革命;(6)一个;(7)没有;(8)问题;(9)就是;(10)这个;(11)战争;(12)工作;(13)这种;(14)一切;(15)现在;(16)发展;(17)但是;(18)群众;(19)主义;(20)这样。

两者重合的二字词汇只有 3 个:我们;一个;这个。去掉这 3 个词汇, 其余 17 个都是各自文本的特色词汇。

《红楼梦》的三字词汇共计 592 种(types), 占总词汇种类 20 004 的 3%; 出现次数总计 8 025 次(occurrences), 占总词汇使用次数 546 327 的 1.5%。词频排列前 20 个三字词汇分别为(不含人名与称谓):(1)怎么样;(2)就是了;(3)为什么;(4)少不得;(5)一句话;(6)了不得;(7)好容易;(8)一点儿;(9)使不得;(10)一辈子;(11)不用说;(12)怨不得;(13)笑嘻嘻;(14)没什么;(15)没意思;(16)点点头;(17)由不得;(18)想不到;(19)一家子;(20)禁不住。

《毛泽东选集》一至五卷的三字词汇共计 1 627 种(types), 占总词汇

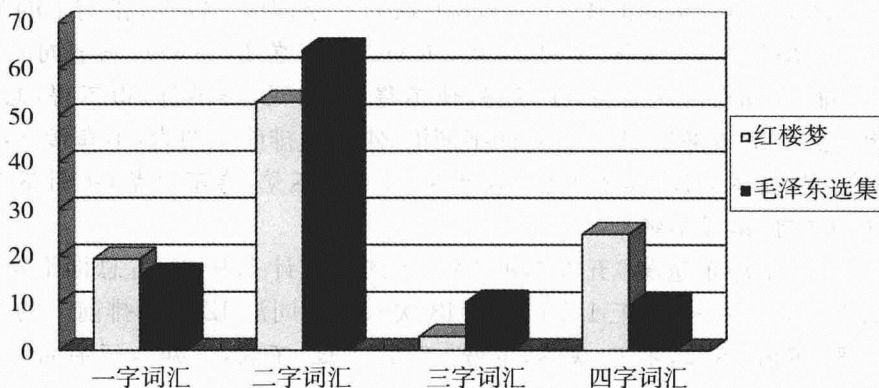
种类 15 755 的 10.3%; 出现次数总计 19 886 次(occurrences), 占总词汇使用次数 552 409 的 3.6%。词频排列前 20 个三字词汇分别为:(1)国民党;(2)共产党;(3)中国人;(4)合作社;(5)根据地;(6)一部分;(7)解放区;(8)反动派;(9)为什么;(10)一方面;(11)解放军;(12)合作化;(13)反革命;(14)共和国;(15)大多数;(16)基本上;(17)侵略者;(18)法西斯;(19)全世界;(20)委员会。

两者重合的三字词汇只有 1 个:为什么。去掉这 1 个词汇,其余 19 个都是各自文本的特色词汇。

《红楼梦》的四字词汇共计 4 938 种(types), 占总词汇种类 20 004 的 24.7%; 出现次数总计 6 657 次(occurrences), 占总词汇使用次数 546 327 的 1.2%。词频排列前 20 个四字词汇分别为(不含人名与称谓):(1)不在话下;(2)不好意思;(3)阿弥陀佛;(4)上上下下;(5)自不必说;(6)鬼鬼祟祟;(7)死去活来;(8)慌慌张张;(9)胡思乱想;(10)呜呜咽咽;(11)无精打彩;(12)大惊小怪;(13)翻来覆去;(14)说说笑笑;(15)放在眼里;(16)感谢不尽;(17)拉拉扯扯;(18)如此这般;(19)喜之不尽;(20)一五一十。

《毛泽东选集》一至五卷的四字词汇共计 1 509 种(types), 占总词汇种类 15 755 的 9.6%; 出现次数总计 14 183 次(occurrences), 占总词汇使用次数 552 409 的 2.6%。词频排列前 20 个四字词汇分别为:(1)资产阶级;(2)社会主义;(3)帝国主义;(4)无产阶级;(5)统一战线;(6)资本主义;(7)知识分子;(8)游击战争;(9)三民主义;(10)抗日战争;(11)日本帝国;(12)土地改革;(13)列宁主义;(14)工人阶级;(15)共产党员;(16)民主革命;(17)半殖民地;(18)共产党人;(19)民主党派;(20)机会主义。

两者重合的四字词汇一个都没有。20 个四字格词汇都是各自文本的特色词汇。



《红楼梦》与《毛泽东选集》五卷本词长百分比对照示意图

在现代汉语中,纯书面体文本二字词汇明显要比综合性文本二字词汇要多,而其一字词汇要少一些。

《红楼梦》里的语体较为中性的否定词“不”的字频排第二,总计15 129个,字频为1.729%;《毛泽东选集》里“不”的字频排第四,总计11 696个,字频为1.177%;《红楼梦》里口语化否定词“没”的字频排第71,总计2 268个,字频为0.259%;《毛泽东选集》里“没”的字频排第94,总计2 095个,字频为0.211%;《红楼梦》里书面化否定词“无”的字频排第109,总计1 401个,字频为0.160%;《毛泽东选集》里“无”的字频排第123,总计1 727个,字频为0.174%。通过“不”、“没”、“无”这三个否定词的统计比较,我们发现《红楼梦》文本尽管比《毛泽东选集》文本早两百年,但是因为其小说文体的缘故,显得更加口语化。

《红楼梦》里“不”的搭配总计739种,排前十位的分别是:不知(1 086)、不是(887)、不得(834)、不好(553)、不过(527)、不敢(502)、不能(435)、不用(309)、不成(267)、不在(258);

《毛泽东选集》里“不”的搭配总计661种,排前十位的分别是:不是(1 281)、不能(1 220)、不要(552)、不可(487)、不但(381)、不同(381)、不会(188)、不应(187)、不得(174)、不好(172)。

《红楼梦》里的“不”字搭配比《毛泽东选集》里的“不”字搭配明显更口语化。两个文本中四个搭配是重合的,两个较为书面体的否定词语“不是”和“不能”无论是排位还是词频,《毛泽东选集》的文本高于《红楼梦》文本,而两个较为口语化的否定词“不得”和“不好”从排位到词频,《红楼梦》文本都比《毛泽东选集》文本要高。

在词频统计中,《红楼梦》的“不”字词汇总计443个,占总词汇量的2.23%,其中一字词汇1个(不6 310次),二字词汇134个(排前十的为:不知,不过,不好,不得,不用,不成,不如,不见,岂不,不必),三字词汇67个(排前十的为:少不得,了不得,使不得,不用说,怨不得,由不得,想不到,禁不住,不得了,舍不得),四字词汇241个(排前十的为:不在话下,不好意思,自不必说,喜之不尽,感谢不尽,不知不觉,喜不自禁,半新不旧,未为不可,络绎不绝)。

而《毛泽东选集》五卷本的“不”字词汇总计329个,占总词汇量的2.09%,其中一字词汇1个(不6 048次),二字词汇127个(排前十的为:不要,不同,不但,不可,决不,不好,不对,不过,不利,不知),三字词汇48个(排前十的为:不得不,差不多,不能不,看不起,不得了,了不起,从来不,不足以,行不通,要不得),四字词汇153个(排前十的为:不可避免,不

切实际,永垂不朽,不可救药,不堪设想,不失时机,不屈不挠,不可收拾,措手不及,毫不犹豫)。

在“不”字的三字词汇中,《红楼梦》里有21种“不得”结构,总词频569次,《毛选》里有11种“不得”结构,总词频31次。《红楼梦》“不得”三字结构的种类是《毛选》的近2倍,词频是《毛选》的18倍。其中《红楼梦》的独特词有13种(303次):使不得(49)、比不得(45)、巴不得(41)、怨不得(40)、由不得(29)、说不得(23)、顾不得(20)、等不得(15)、恨不得(15)、管不得(9)、受不得(8)、免不得(5)、记不得(4);《红楼梦》特色词有6种(262次-12次):少不得(113-1)、了不得(82-7)、舍不得(23-1)、怪不得(18-1)、算不得(17-1)、见不得(9-1);《毛选》独特词4种(5次):经不得(2)、惹不得(2)、迫不得(1)、算不得(1);《毛选》特色词2种(14次-4次):来不得(3-2)、要不得(11-2)。

三字格词汇在文学作品口语体部分出现频率高,而“不得”结构更是口语体词语,基本上在非正规场合使用,在书面体中使用较少。《红楼梦》里569次“不得”结构词汇基本上都出现在比较随意的作者叙述语言中和人物对话里。

曹雪芹与高鹗字频及其搭配比较

经过统计比较,我们发现曹雪芹的字频与高鹗的字频存在较大的区别。曹雪芹文本总汉字数为501 003,总字类为4 100,平均使用次数为122次;高鹗文本总字符为235 019,总字类为3 132,平均使用次数为75次。

《红楼梦》前八十回曹雪芹原作文本列前100位的汉字为:

了;不;的;一;来;道;人;我;是;说;他;这;你;去;个;也;子;玉;有;儿;宝;又;着;们;笑;只;贾;得;见;里;上;便;好;家;在;太;那;姐;大;就;日;出;听;头;下;么;都;知;二;之;要;回;如;两;过;事;自;还;话;心;因;小;老;看;母;起;今;忙;可;时;没;凤;面;些;等;问;娘;奶;吃;此;中;叫;什;呢;才;到;倒;罢;夫;打;想;样;无;方;正;和;后;别;黛;再。

《红楼梦》后四十回高鹗续作文本列前100位的汉字为:

了;的;不;来;道;是;人;一;说;我;这;着;他;里;贾;儿;玉;有;也;宝;你;去;那;太;见;么;只;个;在;们;便;好;得;子;又;家;头;回;上;听;要;大;到;知;老;心;出;叫;姐;爷;起;就;事;话;还;看;过;没;二;些;自;想;都;下;夫;日;王;时;母;呢;问;什;奶;今;才;娘;政;如;进;笑;打;等;姑;凤;怎;两;样;天;罢;几;小;倒;走;外;和;