

# 基于语料库的汉语翻译语体特征多维分析<sup>\*</sup>

西南大学 胡显耀

**提要:** 本文基于汉语翻译和原创类比语料库, 对汉语翻译语体进行多维/ 多特征分析。研究表明: 多维分析法可区分汉语翻译和原创语料, 找出两者具有区别性的语言特征。文中对多个汉语语言特征频率进行了因子分析, 得到两个主要因子。第一个因子可区分文学与非文学语料; 第二因子可区分翻译与原创语料。按因子负荷的高低, 汉语翻译语体的典型特征有: 语法显化程度提高, 助、介、连、代词等均存在显化趋势; 高频词、习用语、成语及汉语特有结构“被”字句和“把”字句等的使用频率显著提高。

**关键词:** 语料库、翻译汉语、多维分析、因子分析、语体特征

[ 中图分类号] H159

[ 文献标识码] A

[ 文章编号] 1000-0429(2010) 06-0451-08

## 1. 翻译语体特征与语料库

20 世纪 80 年代以来, 翻译语言的特征成为翻译研究的一个焦点。从 Toury( 1979) 的“中介语”( inter-language) 到 Frawley( 1984) 的“第三符号”( the third code), 从 Schaffner & Adab( 2001) 的“杂合语言”( hybrid language) 到 Baker( 1993) 的“翻译共性”( translation universals), 再到 Olohan( 2004) 提出的“翻译的特征”( features of translation), 都是将翻译语言作为一种独立的语言变体来研究, 视为翻译本体研究的重点。

翻译作为一种不同语言间的中介行为, 必然与另一语言和文化有关, 原语、原文和原语文化对翻译的影响不可避免。因而, 无论如何归化, 译文总与译入语中的原创文本存在细微差异。这些差异不违背译入语的语法和社会文化规范, 并非“病句”; 也不能以“翻译腔”一言蔽

之, 因为异域情调有时正是读者对译文的期待。造成翻译出现这些特征的因素不仅是原语与译入语的系统差异, 译入语中有关翻译的社会规范也会影响不同时代译文的特征。更重要的是, 翻译文本的特殊性还可能是翻译作为一种特殊的认知行为与生俱来的性质。这种独特性可能具有跨越语言和文化的普遍性或“翻译共性”。

因此, 研究翻译语体特征至少有三个方面的价值: 其一, 通过原语与译入语、原文与译文的语际对比, 了解语言系统差异究竟如何造成了译文的变异, 或者说, 原语怎样通过翻译干扰译入语? 其二, 通过描述和比较同一语言不同时代的翻译语体特征, 我们可判断同一译入语中翻译规范的嬗变( 如不同时期对归化与异化的态度)。其三, 在前两者的基础上, 通过描述和比较不同语言中翻译语体的特征, 我们或可找出跨越语言和文化疆域的“翻译共性”。

<sup>\*</sup> 本文受教育部人文社科规划项目“汉语文学翻译语料库建设研究”( 08JA740036)、王克非主持的北京市优秀博士论文导师专项科研资金项目和西南大学科研项目(SWU 0909404) 的资助。感谢编辑部和匿名审稿人提出的重要意见和建议。

兴起于上世纪 90 年代的“语料库翻译研究”为全面描写一种语言中的翻译语体特征提供了新的方法。利用语料库可对代表某一语体的大量文本的众多语言特征进行提取和统计分析。Baker (1993)、Laviosa (1998)、Kenny (2001) 和 Olohan (2004) 等基于语料库的“翻译共性”研究就是对翻译语言普遍特征的探索。然而, 语料库提供的信息极其丰富, 研究者往往只能从个别信息入手讨论特定的翻译共性, 而不同的信息对同一共性假设常常给出不同甚至互相矛盾的结论。例如, Laviosa (1998) 对翻译英语词汇特征的考察支持“简化”假设, 而 Mairanen (2000) 发现翻译英语存在“异常搭配”更多, 这一证据不支持“简化”。那么, 如何从语料库所提供的大量数据中找出翻译语体区别于原创文本的典型特征呢?

美国学者 Biber (1988/1991) 提出的多维度/多特征分析方法(MD/MF analysis) 为我们提供了一个解决问题的办法。Biber (同上: 73-75) 使用了 67 个语言特征, 运用统计学的因子分析方法描述英语口语和书面语的变体特征。本文在王克非教授主持的北京外国语大学“通用汉英对应语料库”(GCEPC)<sup>1</sup> 的基础上, 用多维分析法来寻找汉语翻译语体的典型特征。

## 2. 汉语翻译语体特征多维分析

### 2.1 特征频率统计

为了分析汉语翻译与原创语料的跨语体(文学和非文学)特征, 通过对 GCEPC 的四个汉语子库的初步分析, 我们采取了 32 个特征作为因子分析的参数(见表 1)。选择这些特征的理由是:

第一、可能区分翻译与非翻译语体的特征。本文选取的特征主要是基于笔者和其他研究者对翻译共性(简化、显化和规范化等)的研究。这些特征大体可分为三类: 1) 是反映语料“简化”(Simplification)倾向的特征, 如: 类符/形符比(TTR), 即语料库中类符(types)占全部形符(tokens)的比例, 用于反映语料库词形的丰富程度<sup>2</sup>; 词汇密度(LD), 指语料库中实词形符

占总形符数的比例, 可作为反映语料库信息量大小和文本难度的指标之一。平均词长(AWL), 即汉语词语的平均字数; 平均句长(ASL)和平均段长(APL), 即汉语句子(或段落)的平均字数, 这三个指标可用于反映语料的可读性(readability)。此外还包括主要实词词类(名、动、形、副、数词等)的频率。词类频率分布可反映翻译语料是否受原语影响出现差异。名词中还单独统计了人名(Nr)的使用频率, 目的在于考察翻译语料的指代功能是否存在差异。2) 是反映语料“显化”(explicitation)倾向的特征, 如语法显化程度(grammatical explicitation, GE), 即虚词比例, 指语料库中各种语法功能词(虚词)占总形符数的比例, 用于反映语料库语法功能词的显化或隐化程度。王克非、胡显耀(2008)发现汉语翻译语料中虚词的使用与原创语料存在显著差异。除虚词整体比例外, 本文还单独统计了助词总频率及使用频率最高的六个助词(结构助词“的、地、得”和动态助词“着、了、过”)和介词、连词、代词等的使用频率。关于翻译语料中助、介、连词和代词(指示代词和人称代词)的特殊表现, 笔者进行了一些相关研究(胡显耀 2007, 2008; 胡显耀、曾佳 2009b), 发现这些语法功能词在翻译语料中存在高频化的趋势。胡显耀、曾佳(2009a)具体考察了结构助词“的”的表现, 发现“的”用作定语标记的频率在翻译小说中显著提高。3) 是反映语料“规范化”(normalization)或“创造性”的特征, 如高频词前十位频率之和(TOP10), 即语料库词表前十位的总频率。语料库中频率最高的词形通常都是虚词或语法功能词, 这些词除提高语料的语法显化程度外, 也表明语料是否过度使用常用词, 降低阅读难度, 提高其在目的语中的规范性和可接受性。因此, 本文用该变量来说明语料的总体规范性。“一次性词”(Hapax Legomena, HPLG)指语料库中只出现过一次的词形, 有学者用这个指标来比较翻译和原创文本的创造性(Kenny 2001)。此外, 我们还选择了一些汉语所特有的语言特征(如被字句、把字句、习用语、简略语、成语等)用于

反映语料的规范性或地道性。综上所述,本文用于多维分析的特征主要来自对翻译共性或翻译语体特征的研究,这些特征重在验证翻译与原创语体的区别性<sup>3</sup>。

第二、提取特征值的技术可能性。限于汉语语料的标注程度,某些非形态化的特征(如汉语意念被动句、无主句等)无法通过形态化的手段提取,故未纳入。因此,本文所统计分析的语

言特征只是语料库技术能够统计的显性特征。具体做法是:统计各特征在四个子库中的原始频率;根据语料库容量以 1000 词为基础计算各个特征的标准化频率;利用统计工具 SPSS13.0 进行因子分析。表 1 为因子分析采用的语言特征标准化频率表。其中,平均词长(字数)、平均句长(词数)、平均段长(词数)采用原始统计值,其他特征均标准化为千分比。

表 1. 汉语因子分析语言特征 标准化频率表(千分比)

序号	特征	代码	A 汉文原	B 汉非文原	C 汉文翻	D 汉非文翻
1	类符形符比	TTR	50.00	49.79	39.86	57.25
2	平均词长	AWL	1.37	1.70	1.44	1.61
3	平均句长	ASL	15.12	19.19	15.86	19.36
4	平均段长	APL	65.44	80.09	51.07	68.31
5	词汇密度	LD	618.13	654.68	571.83	633.76
6	虚词比例	GE	212.80	198.01	264.99	234.38
7	高频词前 10 位	TOP10	179.39	171.78	204.17	170.09
8	一次性词	HPLG	411.23	396.43	352.74	396.98
9	标点符号比例	PUNC	170.04	151.64	158.01	131.90
10	名词	N	247.95	310.09	221.37	298.21
11	人名	Nr	35.26	10.27	16.16	11.74
12	动词	V	206.85	212.99	200.36	202.86
13	形容词	A	46.54	42.15	44.64	41.37
14	副词	D	85.90	54.75	75.70	56.04
15	数词	M	30.89	34.71	29.77	35.28
16	代词	R	74.31	42.48	101.14	53.50
17	指示代词	RZ	17.24	12.41	23.46	13.83
18	人称代词	RP	43.01	17.47	65.11	22.42
19	助词	U	70.70	74.84	83.44	85.15
20	的	DD	47.62	67.87	59.96	78.58
21	地	DI	4.46	3.32	6.20	2.90
22	得	DE	5.82	1.61	4.40	1.43
23	着	ZH	9.07	0.81	8.66	2.03
24	了	LO	27.69	13.13	24.42	11.73
25	过	GO	2.70	1.11	3.89	1.21
26	介词	P	26.06	37.83	37.69	46.19
27	连词	C	17.36	29.24	21.95	33.71
28	被字句	PSV	0.69	1.34	1.37	2.92
29	把字句	BAS	3.31	1.64	4.93	2.44
30	成语	I	2.79	2.92	5.66	2.96
31	简略语	J	3.44	6.38	2.23	3.16
32	习用语	L	1.75	5.41	2.77	2.94

2.2 因子分析

用 SPSS13.0 按特征值> 1 的系统默认值提取因子数量, 对上述 32 个特征的标准化频率进行因子分析, 即可得到各因子的特征值、共享方差以及这些因子所能解释的语言特征及其解释度。上述特征经因子分析后得到 3 个因子, 表 2 为这三个因子的特征值和共享方差的百分比, 可见三个因子能解释的累积方差达到 100%, 说明这三个因子可解释全部语料的特征。

表 2 汉语语料因子分析的特征值和方差

因子	初始特征根		
	合计	提取方差比例%	累计 %
1	21. 35	66. 73	66. 73
2	7. 96	24. 87	91. 60
3	2. 69	8. 40	100. 00

表 3 三个因子所解释的汉语语言特征和因子负荷

第 1 个因子	因子负荷	第 2 个因子	因子负荷	第 3 个因子	因子负荷
过	0. 9786	助词	0. 9804	习用语	0. 7242
着	0. 9764	介词	0. 8122	简略语	0. 5406
人称代词	0. 9656	虚词比例	0. 7957	动词	0. 4547
代词	0. 9551	被字句	0. 6870	成语	0. 3659
了	0. 9363	成语	0. 6470	平均词长	0. 3527
地	0. 9303	的	0. 6393	高频词前 10 位	0. 3122
把字句	0. 9291	连词	0. 4997	地	0. 2966
指示代词	0. 9282	高频词前 10 位	0. 4059	标点符号	0. 2753
副词	0. 9022	把字句	0. 3694	平均段长	0. 1537
得	0. 9012	指示代词	0. 3563	过	0. 1096
高频词前 10 位	0. 8589	平均句长	0. 3030	指示代词	0. 1076
形容词	0. 8584	代词	0. 2961	人称代词	0. 0802
标点符号	0. 7159	人称代词	0. 2472	平均句长	0. 0692
成语	0. 6690	平均词长	0. 2217	名词	0. 0145
人名	0. 6129	地	0. 2156	代词	0. 0135
虚词比例	0. 5927	过	0. 1743		
		数词	0. 0642		
		习用语	0. 0529		

根据所得因子及各语言特征的负荷值, 可画出语言特征的成分图。为便于分析, 我们选取了前两个因子, 画出二维成分示意图(图 1)。该图显示了 32 个汉语语言特征在两个因子(成

表 3 描述的是汉语语料中三个因子所解释的语言特征及其解释度(即因子负荷)。限于篇幅, 只列出了因子负荷为正值的所有语言特征。从中可见, 第一个因子所解释的语言特征主要是时体标记(“着、了、过”)、代词(人称代词和指示代词)、人名、形容词等文学语料的典型语言特征。第二个因子解释的语言特征包括助词、介词、虚词比例, 被字句、成语、结构助词“的”等特征。根据先前的研究, 我们认为这些特征可能是翻译语料的主要特征。第三个因子的解释度(提取比例)有限, 其中, 习用语、简略语、动词、成语等是汉语特有的语言特征, 本文暂不展开讨论。由此可见, 前两个主要因子分别将文学与非文学、翻译与原创语料区别开来。

分或维度<sup>4)</sup>中的分布。首先, 第一个因子(component 1)将所有特征清晰地划分为两部分。右侧负荷值为正, 代表汉语文学语料; 左侧负荷为负值, 代表非文学语料。着、了、过、地、

得等助词、人称代词、高频词等聚集在图的右侧;而名词、动词、数词、TTR、词汇密度等聚集在左侧。其次,第二个因子(component 2)将图1划分为上下两部分,正上方居中为助词——该特征代表了汉语翻译语体最显著的标志。第二个因子上方包括了翻译语料突出的语言特征,如介词、虚词比例、成语、被字句等等。需指出的是,第二个因子上方被分为文学和非文学两部分,右上方表示汉语文学翻译的典型特征;左上方则表示非文学翻译的典型特征。最后,第二个因子下方为汉语原创语料的典型特征。由于本文关注的目标是翻译语料,故这些特征可被视为翻译语体的非典型特征。例如,文学翻译中人名频率较低,标点较少;非文学翻译中一次性词(语料原创性)较少,词汇密度较低等等。

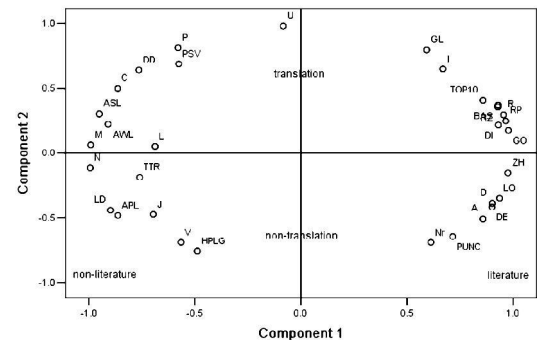


图 1. 汉语语言特征二维成分示意图

综上所述,前两个因子将四个汉语子库的语言特征初步区别开来,这为分析翻译语料的

跨语体特征提供了一个框架,直观地显示了翻译语体的典型和非典型特征。我们看到,这些特征与“显化”、“简化”、“传统化”等翻译共性部分重合。例如,翻译语料中语法功能词/虚词(助词、介词、连词等)词形的增加,表明翻译语料的语法显化程度高于原创语料,或者说,翻译文本存在明确使用语法标记的特征。文学翻译语料中,常用词(TOP10)比原创文学使用更频繁;翻译语料的词汇密度(LD)和类符形符比(TTR)低于原创语料,这两个特征通常被视为翻译文本的简化特征,等等。不过,以上分析建立在我们对因子分析结果的直接观察上,这三个因子是否能将本文使用的四种语料区别开来还需要进一步检验。

2.3 语体辨别

以上对各种语言特征的因子分析共得到了三个因子(成分或维度),并描述了语言特征在前两个维度中的分布。那么,这三个维度是否真能将四个汉语语料库区分开来呢?我们可通过计算各个语料库的“维度分”(dimension score)来加以检验,即根据因子负荷值(见表3),统计各语料库某一因子中特定负荷值(本文取正值)的所有语言特征标准分之和。该方法需将语言特征频率值换算为标准分,计算方式如下。按该方法计算的各语料库语言特征标准分见表4。

$$Z=\frac{X-\bar{X}}{S}$$

(Z 为标准分, X 为原始频率,  $\bar{X}$  为平均值, S 为标准差)。

表 4. 各语料库语言特征标准分

标准分	A	B	C	D	标准分	A	B	C	D
<b>TTR</b>	0.1088	0.0794	- 1.3110	1.1239	<b>RZ</b>	0.1030	- 0.8789	1.3676	- 0.5902
<b>AWL</b>	- 1.0684	1.1301	- 0.5853	0.5367	<b>RP</b>	0.2762	- 0.8978	1.2920	- 0.6703
<b>ASL</b>	- 1.0251	0.8184	- 0.6899	0.8954	<b>U</b>	- 1.1351	- 0.5351	0.7113	0.9591
<b>APL</b>	- 0.0660	1.1620	- 1.2704	0.1746	<b>DD</b>	- 1.2171	0.3340	- 0.2719	1.1544
<b>LD</b>	- 0.0418	0.9968	- 1.3574	0.4024	<b>DI</b>	0.1620	- 0.6106	1.3412	- 0.8952
<b>ML</b>	- 0.5069	- 1.0154	1.2873	0.2350	<b>DE</b>	1.1638	- 0.7926	0.5039	- 0.8762
<b>TOP10</b>	- 0.1249	- 0.6085	1.4498	- 0.7159	<b>ZH</b>	0.9064	- 1.0001	0.8118	- 0.7185
<b>HPLG</b>	0.8636	0.2797	- 1.4440	0.3014	<b>LO</b>	1.0564	- 0.7643	0.6475	- 0.9394

标准分	A	B	C	D	标准分	A	B	C	D
<b>PUNC</b>	1. 0752	- 0. 0791	0. 3205	- 1. 3174	<b>GO</b>	0. 3562	- 0. 8426	1. 2535	- 0. 7672
<b>N</b>	- 0. 5127	0. 9723	- 1. 1480	0. 6884	<b>P</b>	- 1. 3161	0. 1070	0. 0900	1. 1178
<b>Nr</b>	1. 4645	- 0. 7008	- 0. 1905	- 0. 5735	<b>C</b>	- 1. 1227	0. 5028	- 0. 4946	1. 1145
<b>V</b>	0. 1970	1. 3116	- 0. 9812	- 0. 5274	<b>PSV</b>	- 0. 9397	- 0. 2535	- 0. 2218	1. 4145
<b>A</b>	1. 2115	- 0. 6449	0. 4081	- 0. 9747	<b>BAS</b>	0. 1629	- 1. 0215	1. 3119	- 0. 4541
<b>D</b>	1. 1669	- 0. 8749	0. 4983	- 0. 7903	<b>I</b>	- 0. 5717	- 0. 4778	1. 5002	- 0. 4489
<b>M</b>	- 0. 6465	0. 7468	- 1. 0550	0. 9547	<b>J</b>	- 0. 2039	1. 4353	- 0. 8785	- 0. 3600
<b>R</b>	0. 2500	- 0. 9829	1. 2892	- 0. 5561	<b>L</b>	- 0. 9458	1. 4128	- 0. 2885	- 0. 1789

在表 4 的基础上, 我们计算了四个汉语语料库的维度分, 见表 5:

表 5. 各汉语子库的维度分

	语体类型	A 汉文原	B 汉非文原	C 汉文翻	D 汉非文翻
D1	文学与非文学	8. 151542	- 12. 1926	15. 09227	- 11. 0529
D2	翻译与非翻译	- 9. 30956	- 3. 07265	9. 287039	3. 105234
D3	传统与非传统	- 2. 09878	2. 864248	3. 972302	- 4. 73234

表 5 显示了按语言特征, 四个语料库在三个维度上的得分。以下具体分析这三个维度上四个语料库的差别。由表 5 可见, 第一个维度(D1)将文学语料(A 和 C 均为正值)和非文学语料(B 和 D 均为负值)区分开来。具体而言, 在此维度上, A 和 C 的所有特征频率标准分之和高于各自语料库的平均值; 而 B 和 D 的标准分之和高于各自语料库的平均值。值得注意的是: B 汉语非文学原创语料与 D 汉语非文学翻译语料的维度分值相近, 说明二者在所涉及的语言特征上表现相近。而 C 文学翻译语料的维度分(15. 09)几乎是 A 文学原创(8. 15)的 1 倍——这是否意味着文学翻译的“文学性”(由本文所取的特征决定)高于文学原创呢? 由于维度分反映的是这些语言特征的使用频率, 因此, 我们认为汉语文学翻译语料可能的确存在夸大原创文学的某些典型特征的趋势。由于第一个维度主要区别了汉语文学与非文学语料, 因而, 我们将其命名为“文学性”维度。

第二个维度(D2)对本文尤为重要。由表 5 可见, 这个维度将四个语料库区分为翻译(C 和 D 均为正值)与非翻译(A 和 B 均为负值)。从实际分值来看, C 文学翻译的维度分明显高于 D 非文学翻译, 或者说, 文学翻译语料的“翻译

性”(由本文所取特征决定)高于非文学翻译。而 A 文学原创距 C 文学翻译最远, 这说明二者的差异最大; 非文学翻译与非文学原创之间的差异则较小。上述分析说明: 第二个维度所包含的特征具有语体区别性, 基于同样理由, 我们将其命名为“翻译性”维度。

第三个维度(D3)虽然解释力有限, 但也将语料库分为两种类型。从表 3 所列的语言特征来看, 该维度似乎代表了汉语语言特征的“传统性”(成语、简略语、动词等)。换言之, 这个维度所包括的特征大都是汉语特有的语言特征。我们暂且认为它把语料分为传统和非传统两类。有意思的是, 从表 5 的维度分来看, C 文学翻译似乎显得更“传统”(即习用语、成语、简略语、动词等使用频率更高), B 非文学原创次之; D 非文学翻译和 C 文学原创似乎更倾向于打破传统。

2. 4 汉语翻译语体的典型特征

据上述分析, 我们认为汉语翻译语料的确具有区别于原创语料的特征。因子分析的结果显示: 第二个因子可以将汉语翻译与原创语料区别开来。该因子所解释的语言特征(见表 3)及其可能反映的翻译语体特征包括: 1) 显化特征: 如语法显化程度提高(虚词增加), 助词(的、

地)、介词、连词、代词、平均句长等语言特征在汉语翻译语体中均呈现出显著提高的现象;2) 简化特征:如高频词(TOP10)、习用语增多等;三、规范化特征:如被字句、成语、把字句等汉语特有的语法现象使用频率提高等。这些语言特征及其反映的翻译语体共性在其他研究中已得到了一些证实(见胡显耀 2007, 2008; 胡显耀、曾佳 2009a, 2009b, 2010; 王克非、胡显耀 2008, 2010 等)。

### 3. 总结与讨论

以上对文学与非文学的翻译汉语和原创汉语语料进行了多维/多特征的统计分析。统计学因子分析方法的实质是将大量语言现象通过频率值简化为少数几个主要因子(成分或维度),这些主要因子可概括众多语言特征的总体趋势。本文从文体统计学的角度比较系统地验证了这些因子对翻译和原创语料的区别性。我们用计算维度分的方法检验了三个因子是否能区别同一语言中不同的语体。检验的结果说明:不仅汉语文学与非文学语料具有显著差别,汉语翻译与原创语料也存在显著差异。因此,若要明确回答本文提出的问题——汉语翻译语体是否存在?我们的答案无疑是肯定的。翻译文本在某些语言特征上确实存在与原创文本的显著差别,而这些语言特征也就是汉语翻译语体的典型特征。这些特征与目前语料库翻译研究领域讨论较多的显化、简化、规范化等具有相容性。这为我们在不同语言 and 不同语体之间寻找翻译语言的共同特征提供了必要的基础。

多维分析所呈现的汉语翻译语体的具体特征可概括如下:汉语文学和非文学翻译语体与原创相比,存在着语言语法显化程度(虚词比例)提高的趋势,同时,翻译语料的句子长度更长,更集中地使用常用词(高频词)、习用语和被字句、把字句、成语等汉语特有的用法。当然,这些特征在文学翻译和非文学翻译语料中的表现有所不同。总体而言,汉语文学翻译与文学原创的差别大于非文学翻译与原创之间的差异。也就是说,在翻译语体特征表现上,文学

翻译语料的表现更为特殊,而非文学翻译与原创语料比较相近。

上述语言特征中,有些特征(如语法显化程度提高)是趋向于英语(原语)的,这似乎可解释为汉语译文受到语法显化程度更高的英语原语的影响(即原语干扰);而有的特征(如常用词、习用语增多等)趋向于汉语自身,我们暂且用“目的语特征过度呈现”来称呼这种特征。这种特征可视为汉语(目的语)的归化要求。然而,还有一些特征可能是原语干扰和目的语影响无法完全解释的。例如,翻译语料中平均句长增加,定语标记“的”的高频化(及定语结构复合化,见胡显耀、曾佳 2009a)等现象。概而言之,现代汉语翻译语体特征似乎显示出多种力量共同作用的结果:翻译汉语既呈现出语法显化程度提高的原语干扰特征,也存在目的语(汉语)特征的过度呈现,还存在由于翻译行为的特殊认知机制所导致的翻译共性。因此,单独用归化或异化的概念无法说明翻译汉语的总体特征,汉语翻译语体呈现出各种趋势的重叠和杂合。

尽管多维分析法更准确地呈现了汉语翻译语体的一些典型特征,这些特征曾经是传统翻译研究所忽略的,然而,还应看到,其分析结果仅仅是提供了一个“汉语翻译语体”的框架。虽然这个框架中包含了一些具体的语言特征,但由于汉语语料库技术和显性语言特征提取的限制,本文所概括的特征有限。此外,翻译语体特征与语言差异、翻译规范和翻译认知模式的关系还有待分析和解释。翻译语体特征在其他语言(如英语)中是否存在共性?也需要其他语言翻译语体研究的支持。我们将另文论述英语翻译语体特征及其与翻译汉语的共性问题,同时希望此文能引起国内外同行对翻译语体问题的关注。

#### 注 释

1. “通用汉英对应语料库”(GCEPC)包含汉英-英汉双向对应的原创和翻译语料,总容量为 2000 万单词,进行了语料对齐和词性标注。本文使用的语料为 GCEPC 的汉语部分四个语料库的部分语料: A 汉语文学原创语料(46.3 万词), B

- 汉语非文学原创语料(35 万词), C 汉语文学翻译语料(63.4 万词), D 汉语非文学翻译语料(45.5 万词), 共约 190 万词。
- 由于统计汉语标准类符形符比(STTR)不易且精度较差, 而本文所用语料库容量相近, 故使用计算较简便的类符形符比(TTR)。
  - 本文旨在检验这些特征能否区分翻译与原创语料。某些典型汉语特征未入选(如量词、时间副词等), 原因一是初步统计未发现翻译与原创的显著区别, 二是汉语语料库目前检索技术的局限。本文不能也无意于涵括汉语的全部特征。此外, 部分特征间具有覆盖性(如同时收入虚词总频率和各类虚词频率), 本文认为: 二者在不同层面上反映语料特征, 总体频率解释语料整体趋势, 而具体词类则反映语料的实际差异。由于多维分析的目的是找出翻译语体的区别性特征, 无论这些特征是整体还是局部的, 故特征的部分覆盖不影响多维分析的过程。
  - 在因子分析中, 所得结果一般称为“因子”或“成分”。本文 2.2 节讨论因子分析时沿用“因子”, 而论及语体辨别时, 为了强调因子对语体的区别作用, 称“维度”(2.3 节), 二者同义。

#### 参考文献

- Baker, M. 1993. Corpus linguistics and translation studies: Implication and application [A]. In M. Baker, E. Tognini-Bonelli & J. Sinclair (eds.). *Text and Technology: In Honour of John Sinclair* [C]. Amsterdam: John Benjamins. 233-250.
- Biber, D. 1988/1991. *Variation across Speech and Writing* [M]. Cambridge: CUP.
- Frawley, W. 1984. Prolegomenon to a theory of translation [A]. In W. Frawley (ed.). *Translation: Literary, Linguistic, and Philosophical Perspectives* [C]. London & Toronto: Associated University Presses. 159-75.
- Kenny, D. 2001. *Lexis and Creativity in Translation: A Corpus-based Study* [M]. Manchester: St. Jerome Publishing.

- Laviosa, S. 1998. Core patterns of lexical use in a comparable corpus of English narrative prose [J]. *Meta* 43: 474-479.
- Mauranen, A. 2000. Strange strings in translated language: A study on corpora [A]. In M. Olohan (ed.). *Intercultural Faultlines. Research Models in Translation Studies 1: Textual and Cognitive Aspects* [C]. Manchester: St. Jerome. 119-141.
- Olohan, M. 2004. *Introducing Corpora in Translation Studies* [M]. London: Routledge.
- Schäffner, C. & B. Adab. 2001. The idea of the hybrid text in translation: Contact as conflict [J]. *Across Languages and Cultures* 2: 167-180.
- Toury, G. 1979. Interlanguage and its manifestations in translation [J]. *Meta* 24: 223-231.
- 胡显耀, 2007, 基于语料库的汉语翻译小说词语特征研究[J], 《外语教学与研究》(3): 214-220。
- 胡显耀, 2008, 《现代汉语语料库翻译研究》[M]。北京: 外文出版社。
- 胡显耀, 曾佳, 2009a, 汉语翻译小说定语的能力和结构[J], 《解放军外国语学院学报》(3): 61-66。
- 胡显耀, 曾佳, 2009b, 对翻译小说语法标记外显化的语料库研究[J], 《外语研究》(5): 72-79。
- 胡显耀, 曾佳, 2010, 翻译小说被字句的频率、结构和语义韵研究[J], 《外国语》(3): 73-79。
- 王克非、胡显耀, 2008, 基于语料库的翻译汉语词汇特征研究[J], 《中国翻译》(6): 16-21。
- 王克非、胡显耀, 2010, 汉语翻译小说人称代词的外显化和变异[J], 《中国外语》(4): 16-21。

收稿日期: 2010-05-19;

修改稿, 2010-08-31

通讯地址: 400715 重庆市北碚区 西南大学外国语学院





to use break as a transitive verbs, which does not conform to either the verb semantic class hypothesis or the entrenchment hypothesis. In sum, the findings reveal that the L2 acquisition of causative alternation is subject to variables including verb semantics, input frequency and L1 transfer.

**A study of college comprehensive English classroom environment**, by SUN Yunmei (School of Foreign Languages, Huazhong University of Science and Technology, Wuhan 430074, China), p. 438

Using College and University English Classroom Environment Inventory (CUECEI), this study investigated both the students' and teachers' perceptions of their comprehensive English classroom environment in 5 universities. Results of the study indicated that: 1) students' psycho-social perception of their classroom environment is much less favorable than their teachers; 2) students of different genders perceive their classroom environment similarly; 3) students with higher language proficiency perceive their classroom environment more positively; 4) there are certain correlations between students' perception of classroom learning environment and their learning outcomes.

**A contrastive study of lexical proficiency between L1 and L2 compositions via computerized assessment**, by GUI Lin (Institute of English Language and Literature, College of Science and Technology, North Campus, Ningbo University, Ningbo 315212, China), p. 445

The present study investigates lexical meanings and meaning relations in L1 and L2 compositions at the deep-structure level via the computerized writing assessment of Coh-Metrix in order to understand the features of lexical proficiency in L1 and L2 compositions. A quantitative analysis of the six variables that assess lexical proficiency in sample compositions shows significant differences between L1 and L2 compositions and their influence on writing quality. Major differences are as follows: (1) Unlike L1 compositions, L2 compositions use more known information to establish lexical co-referentiality for textual cohesion; (2) L1 compositions manifest more lexical variety and sophistication than L2 compositions through more extensive and abundant meaning networks; (3) L2 compositions are easier to read. Hence, the paper concludes that it is imperative to explore effective measures to cope with such differences.

**A corpus-based multi-dimensional analysis of the stylistic features of translated Chinese**, by HU Xianyao (School of Foreign Languages, Southwest University, Chongqing 400715, China), p. 451

This paper is devoted to a multi-dimensional/multi-feature (MD/MF) analysis of the stylistic features of translated Chinese based on comparable corpora of translated and original Chinese texts. The investigation reveals that the MD/MF analysis is effective in differentiating the translated Chinese texts from the texts originally written in Chinese and in finding out the linguistic features that enable the differentiation. The two factors produced by Factor Analysis can explain most of the linguistic features taken into account. The first factor differentiates the literary from the non-literary texts; the second differentiates the translated from original texts. Based on the factor loads, the following features are found to be the most typical in translated Chinese texts: general tendency towards grammatical explicitation (increased ratio of functional words); higher frequencies of auxiliaries, prepositions and pronouns; higher frequencies of the list-head (most frequent) tokens, idiomatic phrases and some other expressions (like Ba and Bei constructions) that are uniquely Chinese.

**On the influence of mistranslation upon Chinese New-Poem Movement and Anglo-American Imagist Movement**, by WANG Dongfeng (School of Foreign Languages, Sun Yat-sen University, Guangzhou 510275, China), p. 459

It is well known that the Chinese New-Poem Movement and the Anglo-American Imagist Movement are both triggered by translation, the former by the translation of poems from Western countries and the latter by the translation of classic poems from China. This paper probes into the origins of the two movements in question and examines their exponents' translations and their influences. The cases in point reveal that their translations are all full of obvious but well-motivated mistranslations, which share a common agenda, that is, to break into pieces the long-established poetic tradition. Thanks to the misguiding of these mistranslations, the poems of the two movements feature double-foreignization: not only to the poetic formula of the source culture, but also to the poetic tradition of the target culture. Through the (mis)translation, new poetic forms and values are established, and the old local poetic traditions subverted. Interestingly enough, the glories of the two movements are both achieved on the basis of mistranslations.

**Schema-instance hierarchy of metonymy and the mental paths of translation**, by TAN Yesheng (Editorial Board of Journal of Foreign Languages, Shanghai International Studies University, Shanghai 200083, China), p. 465

Within the framework of equivalence theory, research on metonymy as a deviant use can lead to rather limited understanding of the complexity and dynamics of human cognition in translation. Under the cognitive linguistic framework, the paper discusses how the context-based and convention-constrained instantiations of metonymic schemas are related to the shifts and variations in translation. It reveals how the metonymic schema-instance hierarchy based on a diversity of contiguous relations within conceptual frameworks provides the translator with a cognitive creative space. The paper also gives some suggestions on how to further explore into the issues related to metonymy in translation.