

“第三码”再解析——基于自建语料库 CONTEC 的汉译英简化共性量化研究*

凌征华 林泽欣

(江西理工大学 外语外贸学院, 江西 赣州 431000)

摘要: 首先介绍了自建语料库 CONTEC(中国人以英语作为通用语的语料库)的构成,然后利用词汇密度、关键词等分析方法考察词汇特征,利用句法复杂性分析方法探究句法特征,从而检验小说与新闻文本中翻译语言的简化假设是否成立。量化研究结果显示:小说和新闻语类汉译英的词汇密度均高于英语母语,似乎不支持简化假设;而另一方面,汉译英的标准类符/形符比却明显低于英语母语,英语母语用词比汉译英更加丰富,支持简化假设。另外,在句法表层结构上,两种语类的汉译英译文的句子平均长度均高于英语母语,似乎不支持简化假设。而在反映从属性等深层次结构上,新闻语类的汉译英句子的各项复杂性指标却显著低于英语母语,显示出简化趋势,这看似自相矛盾的结果,一定程度上与语料的文体、体裁有关,也反映了翻译共性的细腻和复杂。

关键词: 翻译共性;词汇密度;句法复杂性;语料库;简化假设

中图分类号: H315.9

文献标志码: A

文章编号: 1004-342(2019)04-64-09

语料库翻译研究的一个重要领域就是翻译共性假设及其子假设,这些研究旨在调查翻译语言的普遍特征。最初提出翻译共性概念的是英国学者 Mona Baker。她将其定义为“翻译文本而不是源语文本中出现的典型语言特征,这些特征不是特定语言系统干扰的结果”^[1]。所谓翻译共性是指翻译语言作为一种客观存在的语言变体,相对于源语言或目标语中原创文本从整体上表现出来的一些规律性语言特征,有时候也特指翻译语言本身所固有的特征,被称为“第三语码”或称“翻译腔”^[2]。

自 Baker(1993)提出基于语料库的翻译共性研究以来,围绕该主题已有不少学者采用单语对应语

料库、双语平行语料库的范式进行了大量研究,其中具代表性的有 Baker^{[1][3]}、Laviosa^{[4][5]}、vers^[6]、Olohan & Baker^[7]、Mauranen & Kujamki^[8]、柯飞^[9]、胡显耀^{[10][11]}、吴昂^[12]、Chen^[13]、王克非、肖忠华、戴光荣^[14]、胡开宝、肖忠华^[15]等,内容大体聚焦在翻译的四个共性特征,包括简化、明晰化(显化)、规范化和均质化等。

Gaspari & Bernardini^[16]认为,翻译共性实际上可能属于“中介共性”(mediation universal),因为翻译语言的一些特征也能在非母语文本中找到,两者都是中介语篇(mediated discourse)。因此,他们提出了系统研究中介语(interlanguage)和翻译译文共有特征的

收稿日期:2018-04-20

* 基金基金:江西省社会科学“十一五(2010年)”规划项目(项目编号:10WX43);江西理工大学繁荣哲学社会科学基金项目(项目编号:FZ18-ZD-08)。

作者简介:凌征华(1965-),男,江西理工大学外语外贸学院教授,博士;林泽欣(1992-),男,江西理工大学外语外贸学院硕士。

思路。也就是说,如果在翻译和非母语语篇中发现了相类似的语言特征,我们就可以扩展 Baker 有关翻译共性的假设,认为翻译与非母语产出都受语言接触的影响,因而中介共性(mediation universal)而非翻译共性更能说明这些特征。为此他们利用意大利语使用者的英语笔语语料建了 CONTE(the Corpus of Non-native and Translated English)语料库,以连接词 therefore 为例,对中介共性做了有益的开拓性研究。

考虑到以往的研究主要集中在翻译共性方面,而 Gaspari & Bernardini 虽然开始转向中介共性,但尚未对其作系统的考察,所以本研究自建了 CONTEC (Corpus of Non-native and Translated English by Chinese)语料库,旨在对中介语共性特征进行系统的量化研究,考察汉译英译文(TR)与英语母语(NS),非母语英语(NNW)与英语母语(NS),以及汉译英译文(TR)与非母语英语(NNW)之间,在新闻和小说两种

语类表现出的共性特征。限于篇幅,本文仅对英译文(TR)与英语母语(NS)进行比较,以验证翻译共性特征中的简化假设。

一、CONTEC 语料库

本研究原计划选取 fiction, non-fiction, news 和 academic 四种语类创建语料库,但在创建过程中,由于语料获取困难,最终建成的 CONTEC 语料库由小说和新闻组成,具有“一定的可比性^[17]。”CONTEC 语料库的语料均是公开发表、由中国人直接用英文创作(NNW),以及由中国专家从中文翻译成英文的语料(TR),另外还包括与这两种语类相对应的英语母语参照语料(NS),总容量近 680 万词(见表 1)。语料样本的出版日期基本上在 1991 - 2001 年之间,文本从网上爱问共享资料下载或由纸质出版物扫描,用文字识别软件 ABBYY FineReader 转换成纯文本后,用 CLAWS4 完成词性标注工作。

表 1 CONTEC 语料库选材结构

代码	CONTEC (英文)				
	宏观类型	语 类	词 次	所占比例	语料来源
TR	汉译英	小说 (TR-F)	319 125	4.7%	外研社 1999 版中国文学宝库及爱问共享资料获取的中国专家英译汉语中短篇小说
		新闻 (TR-N)	121 170	1.78%	1994-1998 年宾州中文树库的新闻英译
NNW	汉语使用者 直接用英语创作	小说 (NNW-F)	1 031 403	15.19%	爱问共享资料和百度文库获取的中国移居美英作家作品，日期 2001 年
		新闻 (NNW-N)	2 268 531	33.41%	Gigaword 语料库中随机抽取的 1995-1998 年新华社英文资料
NS	英语母语使用者 英语创作	小说 (NS-F)	485 641	7.15%	FLOB 和 FROWN 两语料库的小说部分
		新闻 (NS-N)	2 563 187	37.75%	Gigaword 语料库中随机抽取的 1995-1998 年美联社 (Associated Press) 英文新闻
总 计			6 789 057		

对于翻译共性特征的检验,以往的研究所采用的测量指标大多只限于平均词长、平均句长、高频词、罕用词、词汇密度,以及标准类符/形符比(STTR)等。后来也有其他学者对测量指标进行了补充,比

如 Williams、胡显耀^[18]利用可读性(Readability Analysis)进行简化和均值化的检验。本研究试图在以往研究成果的基础上,不仅对词汇层面进行考察,而且利用 Lu^[19]的语言复杂性指标来检验翻译共性的句

法特征。

二、汉译英译文呈现的词汇特征

(一)词汇密度

词汇密度的常用测量方式有两种。一种是 Stubbs^[20]定义的实义词数量与总词数之比,用以衡量语篇的信息量。另一种就是语料库语言学中常用的类符/词次比(TTR),即所有类符与所有形符数量之比。由于TTR很容易受到文本长度的影响,所以Scott(2004)提出标准化后的类符与形符比,即标准类符/形符比(STTR),用以更加合理地比较不同长度文本中的词汇差异度。

1. 小说语类汉译英译文(TR-F)与英文母语的词汇密度比较

我们首先采用Stubbs的办法来计算汉译英译文(TR)与英语母语语料库(NS)的词汇密度。结果显示汉译英译文小说部分(TR-F)的平均词汇密度(57.27%)略高于英语母语语料库中的小说部分(NS-F)的平均词汇密度(55.80%),其均差(-1.47)具备统计上的显著性($t = -3.996, df = 31.987, p \approx 0.000$)。结果虽然与Laviosa^[5]和肖忠华^[15]关于英译汉译文词汇密度的结论不一致,即译文的词汇密度略高于母语的词汇密度,却印证了胡显耀的研究结果“英语文学翻译语料的词汇密度还略高于英语文学原创(44.86% > 42.42%)”^[18]。本研究呈现的数据表明两点:一是“从英语的角度而言,来自汉语文学的英语翻译语料在词类分布上趋向于汉语(即实词多虚词少)”的特征(ibid:62)。其二是由于源语和目标语之间的跨语言差异,不同翻译方向(即英译汉与汉译英)会对译文的语言特征产生影响。测量词汇密度的另一种常用方法则是标准类符/形符比。该测量方法主要反映词汇的丰富性和差异度。本研究统计发现TR-F的词汇密度(59.43%)也略高于NS-F的词汇密度(58.74%),但其均差(-0.68158)不具备统计上的显著性($t = -1.277, df = 23.673, p = 0.214$)。结果也与胡显耀^[18]的研究结果相似,而与

Laviosa^[5]和肖忠华^[15]关于译文标准类符/形符比的结论也略微相异,即译文的标准类符/形符比略高于母语的词汇密度,虽然结果不具备统计上的显著性。

2. 新闻语类(TR-N)汉译英译文与英文母语的词汇密度比较

我们同样先用Stubbs的方法计算新闻语类汉译英与母语英文的词汇密度。结果显示汉译英译文的新闻语类(TR-N)词汇密度(62.86%)略高于新闻语类母语英文(NS-N)的词汇密度(61.92%)。其均差(-0.95)不具备统计上的显著性($t = -2.095, df = 10.741, p \approx 0.059$)。

采用标准类符/形符比的方法来测量词汇密度,结果却恰好相反。新闻语类汉译英译文(TR-N)的平均标准类符/形符比(52.19%)明显低于母语英文(NS-N)的平均标准类符/形符比(59.73%),其均差(7.54)具有统计上的显著性($t = 17.163, df = 10.522, p < 0.001$)。与母语英文比较,新闻语类汉译英用词相对贫乏,可能是由于语料库失衡所致。

(二)频率参数

Laviosa^[5]把高频词定义为一个词项出现频率至少占库容0.10%以上的词。在其研究的词表中,有108个词项属于高频词,其中大部分属于功能词。我们在本研究中,也规定高频词所占比例最少为0.10%。

表2 英语母语(NS)和汉译英(TR)中的频率统计

参数类型	小 说		新 闻	
	英语母语	汉译英	英语母语	汉译英
高频词数目	124	133	93	131
累计比例	55.31%	54.37%	43.92%	53.95%
高频词重复率	2 166.27	1 304.55	11 183.62	499.08
高频词与低频词之比	1.237814	1.191492	0.783068	1.171772

表2总结的是英语母语和汉译英译文两个子库的频率参数。可以看出,就小说语类来说,两个子库的高频词数量差不多(124:133),而新闻语类两个子

库的高频词数量则存在一定的差距(93:131)。高频词占英语母语小说语类(NS-F)的比例与汉译英小说语类(TR-F)相差不多(55.31%:54.37%),但高频词占英语母语新闻语类(NS-N)的比例明显低于汉译英译文新闻语类(TR-N)(43.92%:53.95%),这与高频词重复率也相吻合。从表 2 中还可以看出,英语母语新闻语类(NS-N)的高频词重复率(11 183.62)远高于其他子库,很有可能是因为新闻类翻译子库(TR-N)所占总库的比例(1.78%)明显低于英语母语(NS-N)(37.75%)造成的,因为高频词重复率是基于类符(type),而不是基于形符(token)计算的。

接下来分析高频词与罕用词在两个语料库的分布。图 1 给出了频数占所在语料库词次(token)总数大于 0.5%、0.1%、0.07%、0.05%、0.03%、0.02%和 0.01%的高频词的数目。图中可以看到,不论是小说语类还是新闻语类,汉译英译文语料子库中的高频词,普遍都比英语母语的都更常用。也就是说,就词型(type)而言,汉译英译文中的高频词数量多于英语母语。

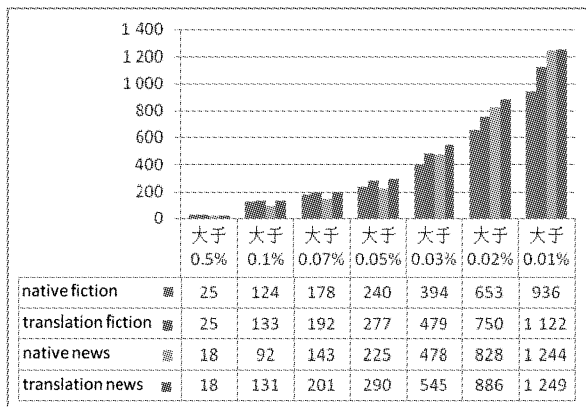


图 1 NS 与 TR 子库中的高频词

(三)关键词分析

关键词分析是语料库语言学中的重要分析手段,本研究中提取汉译英语料子库(TR)中的关键词时采用的参照语料库为对应的英语母语语料子库(NS),因为这样提取出来的关键词针对性特别强^[15]。这一节考察汉译英语料子库中的关键词,包括(正)关键词(即在 TR 中频率显著高于 NS 中频率

的词)和负关键词(即在 TR 中频率显著低于 NS 中频率的词)。

1. 汉译英译文小说语类(TR-F)的关键词分析

TR-F 中最显著的 100 个关键词中,数量最多的是多次复现的名词,尤其是在小说故事中具有中国特色的人名、地名(共计 65 个,如“Tianbao, Liu, mountain, village, Beijing 等),其次是缩略形式(共计 26 个,如 don't, didn't, it's, I'm, I'll, there's 等),还有极少量的形容词、副词和动词等。这些词都是表达实际意义的实义词,它们出现在译文语料库的关键词表上可以说是出于表达内容的需要。^[15]

研究关注的焦点不在比较英语母语与译文的内容,而在翻译体英语的语言形式本身。因此,比上述关键词更有意义的是表 3 中所列的功能词。表中可以看出,汉译英译文与英语母语相比较多使用代词(特别是代词与动词的缩略形式)、连词和连接性的副词,这些词都是语篇中起衔接作用、比较口语化的词。

表 3 汉译英子库 TR-F 关键词

类 别	关键词
P, APPGE(代词)	you'll, his, I'm, it's, I'll, you're, that's, you've, I'd, he'd, we'll
D(限定词类)	this
C(连词类)	after
R(连接性副词)	however

另一方面,汉译英译文语料子库中的负关键词绝大多数是与内容有关的名词,如西方人名 Thomas, Frank, Brown 或与西方习俗有关的其他名词,如 London, lady, Lord, bar 等。但从汉译英译文 TR-F 子库的负关键词中也发现一些有趣的现象,意义模糊、宽泛的副词、名词等,语气较为婉转的情态词用得较少,如 might, maybe, going, toward, almost, stuff。

关键词类分析与关键词分析相似,只不过是基于词类信息而不是具体的词。换句话说,计算关键词类用的是词类的频率而非词频。^[15]表 4 按统计显

著性高低顺序列出了汉译英译文子库 TR-F 中相对于英语母语子库 NS-F 的关键词类和负关键词类。表中可以看出,动词过去分词(VVN)、普通名词(NN)、不定式(VVI)、普通副词(RR)均是高频词类,反映出小说文体特征,描述故事发生的行为和情景。另外,小说对话中“表示程度的副词(RG)用得很多,目的在于使话说得准确,或加重语气,或缓和语气”^[21]。在关键词类排第一的是量词(NNU),则反映出了汉语的迁移影响,因为汉语量词丰富,英语中没有量词,数词可以直接和可数名词连用。

表 4 汉译英语料库 TR-F 中的关键词类

类 型	示 例
关键词类	NNU(度量单位词)、JJ(普通形容词)、VVN(动词过去分词)、VVI(不定式)、RR(普通副词)、APPGE(所有格代词)、CS(从属连词)、II(普通介词)、VVG(动词 ING 分词形式)、MC(基数词)、AT(冠词)、TO(动词不定式符号)、CC(并列连词)、RG(程度副词)、NN(普通名词)、VBDZ(系动词过去式 was)、RP(介词、副词、小品词)、DB(起名词作用的前限定词)、PNQS(主格 wh 代词)、RT(准名词性时间副词)、IW(介词 with, without)、VBDR(系动词 were)、DA(起名词作用的后限定词)、VM(情态助动词)
负关键词类	VBM(系动词 am)、VVGK(going)

2. 汉译英新闻语类(TR-N)的关键词分析

表 5 汉译英子库 TR-N 关键词

类 别	关键词
D(限定词类)	this, such
C(连词类)	and
II(介词)	of, according

在汉译英新闻语类子库(TR-N)中最显著的 100 个关键词中,数量最大的还是译文语料子库中提及的名词(57 个,如 China, Shanghai, economy, science, Xinhua 等),其次是形容词(9 个),动词(7 个),副词

(4 个)。总体上关键词中的功能词用得较少,集中在限定词 this、such、连词 and 和介词 of 等。

表 6 汉译英语料库 TR-N 中的关键词类

类 型	示 例
关键词类	IO(of 介词)、CC(并列连词)、JJ(普通形容词)、NNO(数量词)、RR(普通副词)、RA(名词中心词后的副词)、VHZ(动词 has)、CSN(连词 than)、MD(序数词)、AT(冠词)、DAR(后限定词比较形式)、REX(引导同位结构的副词)、VVG(动词 ing 分词形式)、CST(连词 that)
负关键词类	VVD(动词过去时)、APPGE(所有格代词)、MCMC(连字符连接的数字)、PNQS(主格 wh 代词)、CCB(转折并列连词)、XX(否定词 not, n' t)、VBDZ(系动词过去式 was)、VHD(动词过去时 had)、VBDR(系动词过去式 were)、VVI(不定式 infinitive)、CS(从属连词)、NNB(称呼名词前的词)、VDD(动词过去式 did)、RRQ(wh 普通副词)、VVZ(动词 -S 形式)、PPY(第二人称的人称代词)、DD(限定词)、VHI(动词不定式 have)、MC(基数词)

一般来说,负关键词绝大多数是与内容有关的名词,TR-N 中的负关键词显示的是另一种情况,仅有少量反映当时西方事件的名词,如 Serb, Palestinian, spokesman, war 等。通过其他负关键词,尤其功能词中,可以一些新闻翻译语言的偏向:第一,TR-N 中可能出现少用含有“交互性/信息性”的词:诸如第一、二人称代词,如 we, you 等;否定形式 not, n' t 等;其次,TR-N 可能少用转折并列连词 but, 与 Milton^[22]研究相似;TR-N 还少用从属连接词,如 when, after, where, if, before, until 等;第三,TR-N 中可能出现少用一些介词,如 on, against, about 等)。“交互性”是揭示口笔语之间差别的重要维度,新闻语类子库少用口语词汇,体现出当时新闻语体的特点,比较正式,少有“采访式”原始记录或本语料只是摘译,而少用从

属连词则反映出结构简化的趋势。

再来看一下关键词类分析。表 6 是按统计显著性高低顺序列出了 TR-N 中相对于 NS-N 的关键词类和负关键词类。汉译英新闻语类中,常用词类除了 JJ(普通形容词)和 RR(普通副词)外,最常用关键词类是 IO(of 介词)和 CC(并列连词),反映出汉语对英语的迁移或原语渗透效应,因为“汉语常用短句、散句,通常没有主干和枝杈之分,可以几个小句平行铺排。”“汉语要表达较复杂意义时,……按动作发生的时间顺序和事理的逻辑顺序,逐步交代,层层铺开,一一道来,呈线性递进,在句子结构上不具有层次感,呈平面性”^[23]。“受汉语流水句的影响,中国学生使用英语造句时,往往习惯于写出简单句,然后用逗号或用连词‘and’将其连接起来,并按线性顺序将其排列”。^[21]这种现象也能在后文句法复杂性分析中得到印证,即汉译英译文的并列结构数多于英语母语。在负关键词类中,汉译英子库新闻语类 TR-N 呈现的少用词类是动词的过去形式,包括 VVD(动词过去时)、VBDZ(系动词过去式 was)、VHD(动词过去时 had)、VBDR(系动词过去式 were),反映汉译英中不喜欢使用过去形式报道新闻故事。这种情况很有可能是因为,汉语不存在过去式的曲折形态变化,中国译者在翻译的过程中受到原文的影响。另一少用词类是 PPY(第二人称的人称代词)和 XX(否定词 not, n't),这与以上关键词分析是一致的,即这些词体现了“交互性”,具有强烈的口语特点,在新闻语体少用实属正常。本研究新闻语体中少用转折并列连词 but 现象与 Milton 对香港学生的观察相吻合。

三、汉译英译文呈现的句法特征

对句法特征的考察,以往大多数的研究(Laviosa^[5]、王克非和秦洪武^[22]、McLaughlin、胡开宝)都局限于翻译文本的平均句长、平均句段长、结构容量、简单句和复合句的使用频率等,或者某些典型句子结构的使用频数。因汉英行文差异,句子的长度

不一定说明句子的难度。仅靠平均句长指标并不能有效验证翻译共性假设,其结论也会缺乏生态效度,因此需要更为复杂的综合指标。

自从 Hunt(1965)提出 T 单位概念,并用于评估儿童的语言表现以来,有很多学者对句法复杂度的考察提出了一系列较为有效的测量手段(O'Donnell, et al^[24]、Larsen-Freeman^[25]、Wolfe-Quintero, et al^[26]、Ortega^[27]、Lu^[19]、鲍贵^[20])。比如 Lu^[19]设计了一项第二语言句法复杂性分析软件(L2 Syntactic Complexity Analyzer),建构了 5 类 14 项句法复杂度测量指标,利用 WECCL 语料库数据,对中国英语学习者的句法特征进行了考察,结果显示了非常高的信度。^[19]

第一种类型主要测量子句(MLC)、句子(MLS)和 T 单位(MLT)的平均长度;第二种类型主要包括句子的复杂比率(C/S);第三种类型主要包括反映从属性的四种比率:T 单位复杂性比率(C/T)、复杂 T 单位比率(CT/T)、从句比率(DC/C)和每 T 单位的从句数(DC/T);第四种类型主要是测量并列成分的三种比率,它们是:每个子句的并列词组数(CP/C)、每个 T 单位的并列词组数(CP/T)和每个句子的 T 单位数(T/S);最后一种类型包括三种比率,反映句法结构域更大单位关系的几种比率:每个子句包含的复杂名词词组数(CN/C)、每个 T 单位包含的复杂名词词组数(CN/T)和每个 T 单位包含的动词词组数(VP/T)^[19]。本文试图利用 Lu 的 14 种测量指标对汉译英译文语料子库 TR 和英语母语 NS 进行比较,考察它们的句法特征。

(一)汉译英译文的句法复杂度比较

句法复杂性主要包括单位长度(unit length)和子句密度(clausal density)两个变量。单位长度最常用的 2 个测量指标是 T 单位长度(T-unit length)和子句长度(clausal length);子句密度即句法从属性,最常用的 2 个测量指标是 T 单位复杂性比率(T-unit complexity ratio)和从句比率 T(dependent clause ra-

tio)。^[28]这种分类与 Lu^[19]的十四种类型分类大同小异,但有助于理解 Lu 的复杂性指标。

1. 汉译英小说语类的句法复杂度

表 7 小说语类汉译英译文与英语母语 T 检验

	英语母 语平均	汉译英译 文平均	显著性	均差
类型 1:单位平均长度				
MLC	8.20	9.06	0.001**	-0.86
MLS	12.68	22.06	0.013*	-9.38
MLT	11.53	20.11	0.016*	-8.58
类型 2:句子复杂度				
C/S	1.55	2.39	0.030*	-0.84
类型 3:从属性				
C/T	1.41	2.18	0.033*	-0.77
CT/T	0.32	0.37	0.077	-0.05
DC/C	0.29	0.27	0.015*	0.02
DC/T	0.41	0.57	0.065	-0.16
类型 4:并列性				
CP/C	0.163	0.1591	0.853	0.0039
CP/T	0.228	0.2197	0.787	0.0083
T/S	1.1003	1.1057	0.767	-0.0054
类型 5:特定结构				
CN/C	0.744	0.785	0.351	-0.041
CN/T	1.045	1.686	0.020*	-0.641
VP/T	1.774	2.592	0.018*	-0.818

首先我们采用 Lu^[19]的方法和他为本研究计算的数据,分析小说语类汉译英译文的句法复杂性程度。表 7 显示,从第一、二种类型来看,汉译英译文的子句(MLC)、句子(MLS)和 T 单位(MLT)的平均长度以及句子的复杂度(C/S)均显著高于英语母语相应长度。而从属句和并列句[第三(C/T、CT/T、DC/C、DC/T)、四种(CP/C、CP/T、T/S)类型]的比率在汉译英译文和英语母语两个子库中总体上不存在显著差别,这与对数似然检验值相吻合(LL = -0.015, P = 0.902)。从最后一种类型来看,汉译英译文的 T 单位里所包含的复杂名词词组和动词词组频数

(CN/T、VP/T)高于英语母语。基于以上数据,属于表层的汉译英译文中的单位平均长度和从句数(C/S)均高于英语母语,不支持简化假设。汉译英译文里 T 单位里所包含的复杂名词词组频数高于英语母语显示出规范化的趋势,因为目标语英语倾向多使用名词;而汉译英译文里 T 单位里所包含的复杂动词词组高于英语母语则反映了原语干扰或渗透效应。

(1) 汉译英新闻语类的句法复杂度

表 8 新闻语类汉译英译文与英语母语 T 检验

	英语母 语平均	汉译英译 文平均	显著性	均差
类型 1:单位平均长度				
MLC	11.472	19.271	0.000***	-7.799
MLS	21.811	29.884	0.000***	-8.073
MLT	20.417	29.274	0.000***	-8.857
类型 2:句子复杂度				
C/S	1.902	1.629	0.000***	0.273
类型 3:从属性				
C/T	1.780	1.585	0.000***	0.195
CT/T	0.470	0.374	0.000***	0.096
DC/C	0.356	0.289	0.000***	0.067
DC/T	0.634	0.512	0.000***	0.122
类型 4:并列性				
CP/C	0.232	0.647	0.000***	-0.415
CP/T	0.412	0.973	0.000***	-0.561
T/S	1.068	1.055	0.413	0.013
类型 5:特定结构				
CN/C	1.465	2.652	0.000***	-1.187
CN/T	2.607	4.044	0.000***	-1.437
VP/T	2.344	2.255	0.063	0.089

接下来讨论新闻语类的句法复杂度的各项测量指标。从表 8 中可以看出,汉译英译文的第一类型各项指标的平均长度(MLC,MLS,MLT)均显著高于英语母语的长度。但汉译英译文句子的复杂度(C/S)(第二类型指标)却低于英语母语。第三类的各项

指标(C/T、CT/T、DC/C、DC/T)主要反映从属句比率,如表 8 所示,汉译英译文从属性均显著低于英语母语。但第四类各项指标(CP/C、CP/T、T/S)却显示,汉译英译文使用并列成分的频率明显高于英语母语。类型 5 显示:汉译英译文中的 T 单位名词性词组和动词性词组均高于英语母语。综合这些句子的各项句法复杂性指标可以看到,就句子长度而言,汉译英译文的各项指标高于英语母语;而从句子的深度来看,汉译英译文的各项指标低于英语母语。这说明翻译共性的简化假设不是那么单纯^[15],语言使用不精细化也是简化的表现。

四、结论

本文利用自建语料库对比分析了汉译英译文和英语母语两个对应子库所涉及到的词汇、语法等各种语言特征。通过对汉译英译文和英语母语的比较,发现虽然不同语类之间存在一定的差异性,汉译英译文的词汇密度高于英语母语,似乎不支持简化假设,而更可能受源语的影响,因为汉语是“实词多,虚词少”^[18]。由于源语和目标语之间的跨语言差异,不同翻译方向会对译文的语言特征产生影响。词汇的差异度可以用标准类符/形符比来衡量,汉译英译文和英语母语两个子库的数据说明,汉译英译文的词汇差异低于英语母语,尤其是新闻语体。这就说明原创英语可能比汉译英译文更加多样化。

关键词分析说明,汉译英译文与英语母语相比更多使用代词、动词的缩略形式、连词 and、介词 of 等。这说明与英语母语相比,汉译英译文同等数量的文本传递更少的信息量,是简化的表现,也使语法结构更加明晰化,特别是人称代词和指示代词,因为它们具有衔接上下文的语篇结构。^[15]

负关键词分析显示,小说类汉译英译文少用模糊限制语(hedges)等,如 maybe、almost、might。说明,与英语母语在表达语用稳妥方面,汉译英译者的译文语气需更加细腻。新闻类汉译英译文则少用表示从属性连接词,如 when、after、where、if、before、until

等。

这些均说明翻译共性的简化假设不是那么单纯^[15],语言使用不精细化也是简化的表现。

在关键词类分析中,小说语体的介词 of 和并列连词(CC)的超用很有可能反映了汉译英译文的特色,这也能在句法复杂性分析中得到印证,即汉译英译文的并列结构数多于英语母语。新闻类的度量单位词(NNU),基数词(MC)等在汉译英子库中的超用,则显示原语的干扰或原语渗透效应,因为量词作为汉语的特殊结构在英语中不存在,所以自然会表现在汉译英中。

句法层面的数据表明,在表层结构上,汉译英译文的句子或子句平均长度,并列句使用频率均高于英语母语,似乎不支持简化假设。但在能反映从属性等的深层次结构上,汉译英译文句子的复杂程度却低于英语母语,却又呈现出简化趋势。这进一步说明,翻译共性的简化假设不是那么单纯,语言使用的复杂性应从多个层面考察。

参考文献:

- [1] Baker, M. Corpus Linguistics and Translation Studies: Implications and Applications, in M. Baker, G. Francis and E. Tognini-Bonelli (Eds), Text and Technology: In Honour of John Sinclair [M]. Amsterdam: John Benjamins, 1993: 233 – 250.
- [2] 肖忠华,戴光荣. 寻求“第三语码”——基于汉语译文语料库的翻译共性研究[J]. 外语教学与研究, 2010(1): 52 – 58.
- [3] Baker, M. Corpus-based Translation Studies: The Challenges that Lie Ahead, in H. Somers (Ed.) Terminology, LSP and Translation: Studies in Language Engineering in Honour of Juan C. Sager [M]. Amsterdam: John Benjamins, 1996: 175 – 186.
- [4] Laviosa, S. The Corpus-based Approach: A New Paradigm in Translation Studies[J]. Meta, 1998a(43): 474 – 479.
- [5] Laviosa, S. Core Patterns of Lexical Use in a Comparable Corpus of English Narrative Prose [J]. Meta, 1998b(43): 557 –



- 570.
- [6] Øverås, L. In Search of the Third Code: An Investigation of Norms in Literary Translation[J]. *Meta*, 1998(43): 557 – 570.
- [7] Olohan, M., Baker, M. Reporting that in Translated English: Evidence for Subconscious Processes of Explication? [J]. *Across Languages and Cultures*, 2000(1): 141 – 158.
- [8] Mauranen, A, Kuamki, P. Translation Universals: Do They Exist? [M]. Amsterdam: John Benjamins, 2004.
- [9] 柯飞. 翻译中的隐和显[J]. *外语教学与研究*, 2005(4): 303 – 307.
- [10] 胡显耀. 语料库翻译研究与翻译普遍性[J]. *上海科技翻译*, 2004(4): 47 – 49.
- [11] 胡显耀. 用语料库研究翻译普遍性[J]. *解放军外国语学院学报*, 2005(3): 45 – 48.
- [12] 吴昂, 黄立波. 关于翻译共性的研究[J]. *外语教学与研究*, 2006(5): 296 – 302.
- [13] Chen, W. Explication Through the Use of Connectives in Translated Chinese: A Corpus-based Study [D]. University of Manchester, 2006.
- [14] 肖忠华, 戴光荣. 寻求“第三语码”——基于汉语译文语料库的翻译共性研究[J]. *外语教学与研究*, 2010(1): 52 – 58.
- [15] 肖忠华. 英汉翻译中的汉语译文语料库研究[M]. 上海: 上海交通大学出版社, 2012.
- [16] Gaspari and Bernardini. Comparing Non-native and Translated Language: Monolingual Comparable Corpora with a Twist, In R. Xiao (Ed.) *Using Corpora in Contrastive and Translation Studies* [M]. Newcastle: Cambridge Scholars Publishing, 2010: 215 – 234.
- [17] Laviosa, S. How Comparable can Comparable Corpora [J]. *Target*, 1997(2): 290.
- [18] 胡显耀. 翻译语言特征研究——基于汉英双向平行语料库的翻译共性研究 [PDR]. 北京: 北京外国语大学中国外语教育研究中心, 2010.
- [19] Lu Xiaofei. Automatic Analysis of Syntactic Complexity in Second Language Writing [J]. *International Journal of Corpus Linguistics*, 2010(15): 474 – 496.
- [20] Stubbs, M. Lexical Density: A Computational Technique and Some Findings, in M. Coulter (Ed.) *Talking about Text. Studies Presented to David Brazil on His Retirement* [M]. Birmingham: English Language Research, University of Birmingham, 1986: 27 – 42.
- [21] 王佐良, 丁往道. 英语文体学引论 [M]. 北京: 外语教学与研究出版社, 1987.
- [22] Milton, J. Elements of a Written Interlanguage: A Computational and Corpus-based Study of Institutional Influences on the Acquisition of English by Hong Kong Chinese Students, *Research Reports (V2)* [M]. Hongkong: The Hongkong University of Science and Technology, 2001.
- [23] 杨丰宁. 英汉语言比较与翻译 [M]. 天津: 天津大学出版社, 2006.
- [24] O'Donnell, et al. Syntax of Kindergarten & Elementary School Children: A transformational Analysis, NCTE Research Report 8 Urbana IL [M]. Illinois: National Council of Teachers of English, 1967.
- [25] Larsen-Freeman, D. An ESL Index of Development [J]. *TESOL Quarterly*, 1978(12): 439 – 448.
- [26] Wolfe-Quintero, K., Inagaki, S., Kim, H.-Y. Second Language Development in Writing: Measures of Fluency, Accuracy, and Complexity [M]. Hawaii: University of Hawaii Press, 1998.
- [27] Ortega, L. Syntactic Complexity Measures and Their Relationship to L2 Proficiency: A Research Synthesis of College-level L2 Writing [M]. *Applied Linguistics*, 2003(24): 492 – 518.
- [28] 鲍贵. 英语学习者语言复杂性变化对比研究 [J]. *现代外语*, 2010(2): 167 – 176.

(责任编辑: 刘晓红)

