

# 文学翻译中的语义迁移研究\*

——以基于信息贡献度的主题词提取方法为例

上海外国语大学 胡加圣      上海交通大学 管新潮

**摘 要** 本文尝试提出信息贡献度这一概念,描述翻译文本中的词汇因其对译文特定部分或主题贡献最大的信息量而被赋予相应的权重。信息贡献度有其信息论、语料库语言学、翻译文学、文化传播以及翻译的文化研究等多重理论基础和研究价值。与传统的语料库经典研究方法和大数据研究方法相比,信息贡献度研究方法具有独特作用,可应用于多种与语义相关的研究。借助于 Python 构建的合理算法,实现了文学翻译作品历时性不同译本之间的文化意象的系统化对比研究。研究结果表明,基于信息贡献度方法的语义迁移研究是以文本语义结构作为研究的出发点,不仅保留了传统语料库研究方法的精髓,还为大数据时代 Python 工具在翻译文学和语料库翻译学领域里的应用开拓了新前景。

**关键词** 语义迁移 信息贡献度 Python 文学翻译 主题词 文化意象

**中图分类号** H319.3      **文献标识码** A      **文章编号** 1001-5795(2020)02-0028-0007

翻译学  
研究

## 1 引言

全球化时代离不开文化的交融交流,翻译文学文本的阅读实现了文化的传播和重新认知,在文本翻译和阅读过程中,语义信息会衰减或增强,同一语义符号所承载的文化信息丰富度也相应得到增减,这在同源文本的不同译者、不同译本,尤其是不同时期的不同译文中,体现得更加明显。文化信息处理得当,读者接受的愉悦感越强,翻译文本越会具有“共情”功能,文学翻译才能起到“传情达意”的效果。从这个意义上讲,语言的文化意象的表达是翻译研究值得关注的话题和难点。语言智能技术的发展为解读文化意象提供了有效工具,信息贡献度的提取方法不失为一种新的研究语义迁移的手段。我们甚至可以根据语义贡献的方式、规模及其变化规律,来更好地研究和掌握具有特殊文化意象和文化色彩的词语的翻译和传播技巧,来反推译者的心理、行为、社会、文化、家庭、教育等背景因素,以便更好地解读文本,了解作家作品,促进语料库翻译学和翻译文学的发展。

基于语料库的文学翻译作品的历时研究或共时研究一般都会聚焦于不同译本之间静态的语言

特征对比,其研究路径往往会起始于词汇,终止于理论概念的描述和归纳,如翻译共性、译者风格等,而不同版本的合理时间差恰巧能够为与语言特征相结合的翻译作品的语义迁移研究提供丰富的素材。经典的语料库方法为此提供了诸多工具,如 WordSmith、AntConc 和 ParaConc 等,可将待研究译本与参照语料库(如 BNC)作对比,得出根据关键性排序的主题词,以此对照分析其语义迁移的可能情况。这种参照语料库的方式既有其开拓性优势,但也有其不足之处,即可能没有可参照语料库,或可参照语料库的代表性不充分。为了克服这种因静态分析而可能产生的缺陷,有必要启用更为灵活多变的研究技术;语料库研究中的技术创新与理论深化有着同等重要的作用(许家金,2019;秦洪武等,2014;卢植、胡健,2018)。

本文试图以基于信息贡献度的主题词提取的方式作为经典研究方法的补充,对文学翻译作品的历时性语义迁移展开实证研究,以为语言文化意象的智能化建模研究提供坚实的方法论基础,同时为大数据时代语料库翻译学的创新研究贡献更多有效的工具研发切入点。

★ 版权所有  
文责自负 ★

## 2 研究方法描述

本节通过对比参照语料库方法和词向量模型训练方法,对信息贡献度的定义、作用及其应用条件进行描述,并指出信息贡献度方法和词向量训练方式均已考虑到文本词汇的语义结构,而参照语料库方法仅为独立看待具体的词汇。

### 2.1 信息贡献度

书面语是一种以特定词序传递信息的交际符号。在受到语法约束的同时,文本的语义与话题结构也会对用词模式产生影响。文本中的任何一个单词都会为文本的特定部分或主题贡献其最大的信息量。对于一个给定词频的单词,其在随机文本中随机共现的非均质分布越显著,对整体信息的贡献也就越大(Montemurro & Zanette, 2010)。因此,信息贡献度是指文本中的单词因其对文本特定部分贡献最大信息量而被赋予的权重比例。之所以定义为信息贡献度,其理由在于它不仅是一个语义概念(王勇等, 2018),也是基于信息论而实现的一个概念。

若将一个文本分割成具有最优规模大小的不同部分,其所包含的单词分布将在统计学上构成最大的多样性(Montemurro & Zanette, 2010)。这是计算信息贡献度的统计学依据,即最优规模的确定将直接影响到根据信息贡献度提取主题词的结果。根据 Montemurro & Zanette 对达尔文《物种起源》、罗素《心的分析》和梅尔维尔《白鲸》等多种文本(皆为英文文本)的统计研究,适用于文学或学术文本的最优规模介于 1,000 至 3,000 词之间。由此可推断出适宜于确定信息贡献度的最小书面语文本长度应该是二三千个单词。根据单词的信息贡献度进行排序,最富信息的单词就是文本中那些具有明显的信息代表性的单词。此类最富信息的单词可运用本文的信息论方式加以识别,而无需任何有关语言或文本的先验知识(Montemurro & Zanette, 2010)。根据我们对金融类英汉文本的测试结果,发现英文类的最小文本约需 2,000 词,中文类约为 4,000 字。对照《中华人民共和国著作权法》的中文文本及其英译本(分别为 8,553 字、5,966 词),汉英两种版本均可提取出高质量的几乎完全对等的主题词。按信息贡献度提取主题词的技术可通过 Python 加载 gensim 工具包实现。

信息贡献度的出现是对参照语料库方法的有力补充,弥补了在无法获取具有代表性的参照语料库情况下所显现的不足。

参照语料库方法是指把一个文本或其词频表与参照语料库文本或词频表进行对比,通过卡方检验等方式,提取出按关键性排序的主题词(Scott, 1997)。这一方法可通过 WordSmith 的 Key words 功能或 AntConc 的 Keyword List 功能实现。参照语料库方法已被诸多应用的原因在于,当需要与权威语料库如英国国家语料库(BNC)等进行对比时,其具有显而易见的权威性和标杆性。

而信息贡献度方法是对长距离语义结构进行信息论量化处理的方法,即信息本身是由与具体单词相关的信息分布所构成,最优规模所表征的是文本划分之上的单词分布,与以同样的词频但又随整个文本而随机出现的单词分布有着本质区别(Montemurro & Zanette, 2010)。就此意义而言,信息贡献度方法与参照语料库方法之间的本质区别是,前者考虑了文本词汇的语义结构,而后者仅涉及独立词汇。因此,按信息贡献度提取的主题词,除了按参照语料库方法的应用形式外,还可应用于与语义相关的研究,如语义网络分析、语义迁移研究等领域。

### 2.2 语义迁移研究方法对比

传统经典的语义迁移研究主要通过从历史文本中搜索目标词,统计词汇的使用频次,根据语言和历史知识对其进行人工描述(陆晓蕾、王凡柯, 2020)。翻译研究领域的语义迁移研究,如 place 不同形变的概念范畴语义迁移现象,说明文学作品翻译的语义连贯和情景的可体验性要求译者通过范畴进行跨域语义迁移,以合理表达出其语境化内涵(张伟华、曾利沙, 2014)。又如译者对原语意义的理解、迁移均受认知规律的指导,在翻译中译者根据现实需要,必须有针对性地构建文化语境,因此翻译过程中的语义迁移并不是简单的移植过程,而是基于语境的再创造(王卫强, 2010)。二语习得领域的语义迁移研究,如以语料库方法对中国 EFL 学习者使用同译词 refuse/reject 的频率、搭配行为、类连接进行考察后发现,中国 EFL 学习者并没有完全掌握两个英语同译词的区别,这种不完全习得可能是由于共享的母语对译词占据了两个英语同译词的词条信息,使得两个词条发生重叠和融合所致(王庆、潘雪瑶, 2018)。语义演变中的英语副词 a-gain 的语义发展过程说明,“词汇内在的系统调整和语言使用者在特定语境下的认识都会影响词汇语义的变化,即语义的动态发展既有范畴属性变化的原因,也有语言使用语境的原因。范畴属性的流动性是动态范畴

化的客观原因,为范畴的历时构建提供可能性;而认知主体的概念识解是语义演变的主观原因,为词义中的语法功能和语用意义构建提供可能”(林璐、王旭,2019)。上述研究表明语义变化涉及横向的语义迁移和纵向的语义演变,而本文的研究涉及两者,即翻译中的语义迁移和语义在不同时期译本中的历时变化。为统一起见,均使用语义迁移一词,即横向语义迁移和纵向语义迁移。语义迁移研究也需要更多创新,如构建的新理论框架,从深度和广度两个方面进一步拓展语义迁移的基本证据(姜孟、邬德平,2013)。

大数据时代下的语义迁移研究已显现出研究方法的独特性,即以大数据为基础,通过训练词向量模型,对所获取的词语搭配信息进行考察,进而得出其所反映的社会文化变迁方面的知识。作为表达语义的基本单位之一,词是自然语言处理的主要对象,而词向量的本质就是将人类符号化的词进行数值或向量化表征,以便于计算机的读取、计算和输出(陆晓蕾、王凡柯,2020)。词向量模型的训练需要巨量的语料文本,如刘知远等(2016)借助1950年至2003年的《人民日报》文本训练词向量模型,Hamilton *et al.* (2016)使用了上万亿形符的大型历时多语种语料库。他们训练词向量模型的目的都是为了研究词汇的语义迁移问题,即通过观测词汇的相邻词语分布随时间变化情况分析词汇语义迁移和社会变迁问题。训练词向量模型的研究意义是显而易见的,但模型的训练似乎必须满足两项要求,一是巨量的语料,二是与此对应的高配置计算设备和越来越复杂的算法。对语料数量无法满足词向量模型训练要求的特定文本的语义迁移研究而言,可能仅仅具有象征意义,无法通过词向量来体现大数据特征。

传统经典的研究方法的关键在于目标词的搜索和确定,而大数据研究方法的关键是大数据的获取和词向量模型的训练。前者是在后续解读目标词的过程中完成语义建构,而后者是通过词向量模型的训练来建构语义关系并从中获取语义解读信息。本文所采用的信息贡献度方法则是从文本词汇的语义结构出发来解读词汇的语义迁移现象,可谓兼具两者的特性,又避免了巨量数据的种种限制(见表1)。

### 3 数据信息的获取途径

本研究获取数据信息的过程涉及在Python中加载下述工具包:gensim、jieba、nltk、pandas、re、zhon、

表1 三种不同研究方式对比

方法	目标词	语料库要求	实现技术
传统经典方法	无语义关系 人为设置	参照语料库对比	传统语料库工具 编程技术
信息贡献度方法	根据语义关系权重 选定	语料库自身对比	编程技术
词向量训练方法	根据训练结果所显示 的语义关系选定	超大型库容的 语料库	编程技术

scipy等,用于处理文本语义,以及加载中文分词处理模块和频率分布处理模块等。数据处理过程中使用特定词表进行了数据清洗。上述多种工具包相辅相成,构成获取数据信息过程中不可或缺的一个环节。

本文获取数据信息所涉算法的流程优化步骤(管新潮、熊秋平,2012)如下:

第一步:文本加载、清洗,按信息贡献度分别提取不同版本的主题词;

第二步:提取不同译文共现主题词;

第三步:从共现主题词词表中提取具有鲜明文化含义的主题词;

第四步:分别提取不同主题词的“左五右五”搭配;

第五步:对“左五右五”搭配进行次级主题词确认;

第六步:按第三步主题词提取英汉对应句对,以备解读之需;

第七步:过程数值检验。

第四步的“左五右五”搭配是语料库语言学常用方法,同时也关注到有的研究所采用的是跨距为-5/+5的搭配设置(王庆、潘雪瑶,2018)以及词向量模型训练中多采用类似的窗口设置(陆晓蕾、王凡柯,2020)。虽然方法各异,但对相邻搭配的观点是一致的,因此本文考虑到研究对象的数据量问题也采用“左五右五”搭配设置。

用于获取数据信息的文本是林语堂的《京华烟云》张振玉译本(1977年出版,字数为527,717,以下称“张译本”)和郁飞译本(1991年出版,字数为479,362,以下称“郁译本”)以及英文版原文,两个版本时间相差15年。林语堂是一位双语作家,由于他的母语是汉语,其代表作英文作品《京华烟云》中的语言文化意象因而具备了地道的中国文化特色,并使之具有最标准的双语互译转换识别和自动转换的使用价值。两个译



本也是代表性译者的畅销作品,其出版时间的不同和出版地域的差异性为语义迁移解读提供了诸多可能,尤其是两个译本近 5 万字的汉语译入语差值,也可为界定和分析文学翻译作品的语言特征、研究译者行为和翻译的创造性叛逆等(谢天振,2012)提供巨大的技术想象空间。

## 4 文本数据分析

### 4.1 文本数据概述

从获取的文本数据信息计量,按信息贡献度提取主题词,张译本为 5 321 个、郁译本为 5 375 个,两者的数字差异绝对值并不大,但两译本共现主题词仅为 2 887 个,分别占各自译本的 54.26% 和 53.71%,均为半数以上。也就是说,两译本的主题词类符数差异不大,但所选用于表达原文的同译词(王庆、潘雪瑶,2018)可能存在显著的差异性。再加上前文述及两译本近 5 万字的字符差值,表明各自的形符数存在较大差异。将上述译本字数和非共现主题词进行卡方检验( $p\text{-value}=0.0000$ )表明,两译本在译文用词方面存在显著差异,故可确定,将两译本用于词汇语义迁移研究有其充分的理据。字符差异性和主题词差别不大这两个事实同时也说明,语义迁移研究与总字符数几无关系。其缘由在于不管谁是译者,译文的文化内核基本上是稳定的,因此选取这两个译本作为语义迁移研究的对象也具有翻译文化研究的案例价值。

为突出研究重点,本文将聚焦于林语堂作品汉译文本中多以写“人”为主的词汇,这些称谓在中国文化里具有极强的文化色彩和社会、身份表征。我们将两个译本中描写“人”的词汇大致划分为七类:男人、女人、普通人、特殊人、老人、文人、外国人。下文将选择“男人/女人”和“普通人/特殊人”这两个词对中的具体称谓作为案例进行说明。

### 4.2 男人/女人词对

以男人/女人词对两译本中的“大少爷”和“狐狸精”两个具体词语为例。从信息贡献度的比值看,张译本的“大少爷”是郁译本的 2.38 倍,而郁译本的“狐狸精”则是张译本的 3.35 倍。这一对表示男人/女人的用词在两译本中的信息贡献度权重有较为显著的差异性,张译本的“大少爷”信息贡献度明显大于郁译本的,而“狐狸精”一词正好相反。再看两词的搭配主题词情况(见图 1 和图 2):张译本的“大少爷”前十个搭配主题词为“我、她、牛家、大少爷、没有、照顾、邀请、

姨太太、不、不敢”,郁译本的为“她、我、你、差不多、从小、伺候、难道、吃掉、不得不、牛家”;张译本的“狐狸精”前十个搭配主题词主要为“我、你、太太、她、知道、全家、一定、离开、家事、您”,郁译本的为“我、太太、哪里、一家子、离开、媳妇、当家的、骂、我们、迷住”。

	A	B	C
1	。 经重 和 牛家 的	大少爷	与 男人 同桌, 郭
2	不敢 哼 一声 儿。	大少爷	娶 了一个 愚蠢 软弱
3	两位 待 我 不错,	大少爷	心肠 很好。 这
4	了 大少爷? 我 把	大少爷	吃 了吗? ”
5	什么 东西。 体仁 这位	大少爷	说, “ 我 不
6	低不就。 她 不 能 等到	大少爷	成家。 可是 她 在
7	先来, 她 有 福气 伺候	大少爷	。 她 已经 二十多。
8	即使 银屏 要 走,	大少爷	还 不 一定 肯 放 呢
9	呢? 我 听说 牛家	大少爷	专门 诱拐 人家 的 姨太太
10	孩子 是 我的。	大少爷	跟 我 一点 儿 也 没有
11	我 也 没有 吃 了	大少爷	? 我 把 大少爷 吃
12	我愿 立刻 受 处罚。	大少爷	是 出 洋 念 书 去 了
13	注意。 散戏 之后,	大少爷	的 保镖 就 “ 邀请
14	照顾 躺 在 地上 的	大少爷	。 那个 灯笼, 却
15	的 关系。 牛家 的	大少爷	才 在 政府 一个 小
16	燕 包子, 她 心想	大少爷	会 自己 照顾 自己。
17	还 看 得出 富 贵 之 家 的	大少爷	。 跟 街 上 的 穷 骨 头
18	邀请 ” 那位 姨太太 到	大少爷	的 私 邸 去 过 夜。

(上图)

	A	B	C
1	“ 她 差不多 从小 侍候	大少爷	的, 早起 给 他
2	“ 难道 我 吃 掉了	大少爷	不成? ” 狂怒 的
3	为了 你, 不 得 不 把	大少爷	送 出 洋, 你 使
4	呢? 我 听说 牛家	大少爷	专门 拐 骗 人家 姨太太 来 供
5	就是 她 愿意 走,	大少爷	也 未 必 肯。 她
6	待 我 很 好,	大少爷	更 是 心 善, 我 这
7	这个 家门, 有 福气 伺候	大少爷	。 可是 她 已经 二十 出 头

(下图)

图 1 两译本“大少爷”一词的搭配(上图张译本,下图郁译本)

	A	B	C
1	。 你 知道 那个 小	狐狸精	, 现在 家事 都 由
2	。 太太 以 为 我 是	狐狸精	。 您 两位 待 我
3	全家 都 被 她 这个	狐狸精	迷 住 了。 若 不 是
4	她 说 二 太太 是	狐狸精	。 ” 丁 妈 说:
5	我 一 定 要 离 开 那 个	狐狸精	。 你 亲 眼 看 见 了,
6	是 不 是 真。 我 是 不 是	狐狸精	, 与 你 没 有 关 系

(上图)

	A	B	C
1	。 一 家 子 都 让 那 个	狐狸精	迷 住 了, 要 不 是 有
2	。 太太 当 我 是	狐狸精	, 可 是 你 们 待 我
3	只 要 能 离 开 那 只	狐狸精	, 我 怎 么 都 行。
4	媳 妇 在 哪 里? 那 个	狐狸精	在 哪 里? 我 这 条
5	是 真 是 假。 我 是 不 是	狐狸精	同 你 什 么 相 干,
6	真 正 当 家 的 是 那 个 小	狐狸精	。 ” 莺 莺 明 白
7	骂 我 们 二 太 太 是	狐狸精	。 ” 丁 妈 说:

(下图)

图 2 两译本“狐狸精”一词的搭配(上图张译本,下图郁译本)

	A	B
1	Although my old eyes are dim, I can still recognize a gentleman's son from a poor stuffed bag in the streets.	虽然我老眼昏花,还看得出富贵之家的 <b>大少爷</b> ,跟街上的穷骨头不一样。
2	And, even if she were willing to go away, he might not agree.	并且,即使银屏要走, <b>大少爷</b> 还不一定肯放呢。
3	At least once it was established that a certain person's concubine attracted the <b>elder son's</b> attention and after the theater was "invited" by his bodyguard to one of his private establishments for the night.	至少,有一次,是千真万确,某人的妾惹起牛家少爷的注意。散戏之后, <b>大少爷</b> 的保镖就“邀请”那位姨太太到 <b>大少爷</b> 的私邸去过夜。
4	But even with the best of intentions, Yuan found it impossible to help him. It was some years later that through the influence of Suyun's husband, the <b>elder brother</b> got a minor job in a government bureau.	过了几年,由于素云的丈夫经亚的关系,牛家的 <b>大少爷</b> 才在政府一个小机构里,弄到一个低级员司的差使。
5	Chinya and the <b>elder New boy</b> sat at the men's table, while Sunya and the younger New boy sat at a separate table with the four girls.	经亚和牛家的 <b>大少爷</b> 与男人同桌,荪亚和牛家的小男孩子则和四个女孩子一同坐。
6	I hear that the <b>elder young New</b> specializes in abducting people's concubines for his own purposes.	我听说牛家 <b>大少爷</b> 专门诱拐人家的姨太太呢。”
7	Laima was helping in the kitchen making New Year puddings, and she thought the <b>young master</b> could take care of himself.	赖妈在厨房帮着蒸包子,她心想 <b>大少爷</b> 会自己照顾自己。
8	Lota had hastened to put on his jacket, and came out into the dark courtyard, barely lighted by the flickering oil lamp which Lotung had hurriedly set on the floor while he attended to the <b>young master</b> .	罗大赶紧披上棉袄,跑到这个黑院子里来,这时院子里只有罗东,匆匆忙忙点着那个摇晃不定的灯,正忙着照顾躺在地上的 <b>大少爷</b> 。
9	She cannot wait until he is married, and she is too used to the comforts of this family to go home and marry a country lad, and she is unwilling to leave Peking.	她不能等到 <b>大少爷</b> 成家。可是她在姚家舒服日子过惯了,没法子再去嫁个庄稼汉,并且她也不愿离开北京。
10	Stung by shame, Silverscreen wept aloud, covering her face with her arm and answering, "I have not eaten up the <b>young master</b> , have I?"	羞辱的话伤人太重,银屏号啕大哭起来,用一只胳膊挡着脸说:“我也没有吃了 <b>大少爷</b> ?我把 <b>大少爷</b> 吃了吗?”
	This <b>elder son</b> was married to a weak, obedient, and stupid girl,	<b>大少爷</b> 娶了一个愚蠢软弱倒是百依百顺的女子,做梦也没有梦到过问丈夫到

图3 张译本“大少爷”双语对应部分句对

分析两译本“大少爷”一词通过搭配而突显的语义信息看出:张译本的“大少爷”性格特征似乎为唯唯诺诺、生怕有不当之处,而郁译本所体现的性格特征则显得有些得过且过、无自主性特点。“狐狸精”一词在张译本中体现出自以为是、独断专行的特点,而在郁译本中则是得理不饶人、目中无人的特点。以纵向语义迁移视角看,从张译本到郁译本,“大少爷”的比值权重已有显著下降,两者所体现的性格特征有明显的差异,是从“小心谨慎”过渡到“人生的无所谓”;“狐狸精”一词则权重显著增加,表现为两者性格特征的显著差异性,是从“孤芳自赏”过渡到“出言不逊”。

在双语平行句对中检索“大少爷”一词(见图3)发现,其张译本对应的英文原文有8个词/组(分别出现在16个句对中),即 brother、he、son、elder New boy、elder young New、elder son、young master、eldest young master;郁译本对应的仅为5个英文词/组(分别出现在7个句对中),其中原文的 son、elder New boy、elder son 并未译成“大少爷”。句对统计结果显示,张译本出现18次“大少爷”,而郁译本仅出现7次,所对应的原文为18次,“大少爷”一词并非与原文呈一一对应关系。从原文到译文,两译本“大少爷”一词已显

现出不同的分布强度(见图4)。根据这一组汉英对照数据,从横向语义迁移视角看,张译本更注重“大少爷”一词的应用,这与前述的信息贡献度为2.38倍比值相呼应。

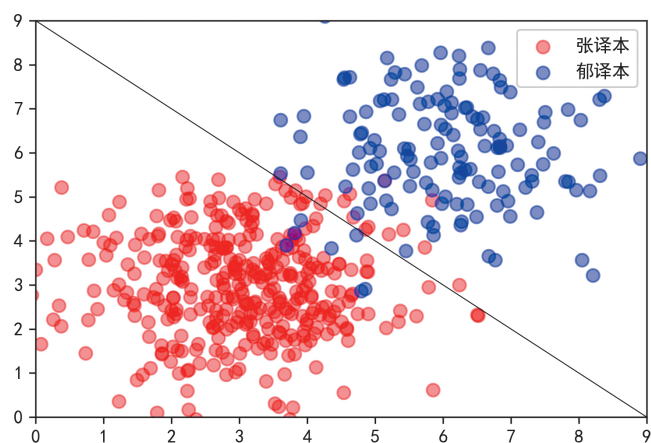


图4 “大少爷”一词两译本分布强度对比

#### 4.3 普通人/特殊人词对

以普通人/特殊人词对中的“正人君子”和“老百姓”两词为例,从信息贡献度比值看,张译本的“正人君子”是郁译本的0.87倍,而“老百姓”则是郁译本的6.12倍。这一对表示普通人/特殊人的用词在两译本



中的信息贡献度权重显示,一方面是差异性较大,而另一方面是差异性不大,即张译本的“老百姓”信息贡献度明显大于郁译本的,而“正人君子”一词则接近等值。再看两词的搭配主题词情况(见图 5 和图 6):张译本的“正人君子”前十个搭配主题词为“周刊、一派、作家、好、朋友、想象、民众、运动、奸诈、邪恶”,郁译本的为“作家、想得到、敌视、民众、那群、讽刺、运动、当局、责怪、学生”;张译本的“老百姓”主要为“普通、她、中国、没有、因为、善良、怕、搜刮、生活、联合”,郁译本的为“我、没有、搜括、她、全国、无不、捧腹大笑、北平、一天、很迟”。

	A	B	C
1	一个周刊,是“	正人君子	派”办的,
2	另一派作家,“	正人君子	派”,当局认为
3	好朋友,是个	正人君子	。”“不行。
4	想象不?那些“	正人君子	”还把过错推
5	民众运动。这群“	正人君子	”大多是英美大学
6	的好诈恶毒而貌似	正人君子	。正如怀瑜那样。
7	那个司令官并不是个	正人君子	——您知道进关
8	,可以说是一个	正人君子	。自律严,有

(上图)

	A	B	C
1	一派作家,所谓“	正人君子	”,当局认为不是
2	想得到吗?那些“	正人君子	”责怪的是学生
3	敌视民众。那群“	正人君子	”的喉舌某周刊
4	而以讽刺这些“	正人君子	”为最大的乐趣
5	这个运动。这些“	正人君子	”多数是英美留学生

(下图)

图 5 两译本“正人君子”一词的搭配(上图为张译本,下图为郁译本)

两译本“正人君子”一词的搭配语义信息与正人君子本来的语义特性颇为接近,均表现为责任、担当的意思。“老百姓”一词虽然信息贡献度差值较大,但通过搭配而突显的语义信息与原义也较为接近,均表现为普通、善良的特性。但从横向语义迁移视角看,通过比较双语句对,发现张译本“老百姓”一词分别出现在 36 个句对中,而郁译本仅出现于 4 个句对,这一点与信息贡献度比值相呼应,说明张译本对“老百姓”一词的选用与郁译本形成了鲜明的对比(参见图 7 的“老百姓”一词对比分布强度)。又从纵向语义迁移视角看,从张译本到郁译本,“正人君子”实现了较为平稳的过渡,语义迁移现象并未显著发生;而“老百姓”一词虽然搭配相近似,但从张译本异常强调语篇氛围过渡到语篇氛围的平平淡淡,表明语义迁移的方向是弱化和淡化。

	A	B	C
1	、抢劫、奸淫。	老百姓	对她感激万分,虽然
2	。他们成了平常	老百姓	,回家各奔营生。
3	。因为真正的中国	老百姓	是扎根在中国的
4	不是登大雅之堂的普通	老百姓	的打算,你乘
5	不过三十里地之远,	老百姓	惶惶不安,觉得大难即将来临
6	人说:“咱们	老百姓	在这儿说没有用
7	今年土地爷生日晚,	老百姓	是欢天喜地。整神之后
8	似的玩意儿,向	老百姓	买东西!简直做
9	作家会以为,是普通	老百姓	写的呢。”
10	做官儿的,欺压	老百姓	的?你若因为
11	厉害。这村子的	老百姓	帮忙很大,我虽
12	各处乱抓青良的	老百姓	吗?”陈三的
13	合心意;另一类普通	老百姓	、保国抗战的士兵
14	大吵起来,北平的	老百姓	就是这样儿,甚至一个
15	她救了好多中国	老百姓	,使他们免于外国
16	她是女人,但是	老百姓	以“董二爷”
17	常常野望做个普通的	老百姓	,现在她的理想
18	开会、游行示威。可是	老百姓	被压制之下日趋高涨
19	报上登出来,怕激起	老百姓	的反日情绪,没法儿
20	搜刮到几千万铜元,	老百姓	是很少知道的。
21	操持,我没有搜刮	老百姓	,拿的只是做官
22	正常生活。最初,	老百姓	并不很怕占领
23	此日情况,在普通	老百姓	之间,汉人并不
24	现在都已开放供	老百姓	游逛。她去游
25	的很多。告诉后方	老百姓	给我们送御带。

(上图:部分截图)

	A	B	C
1	。我没有搜括过	老百姓	,只是救下以为官
2	她:“要是全国	老百姓	都像您啊。
3	无不捧腹大笑。在北平	老百姓	这是常有的事。
4	这一天来得很迟,	老百姓	兴高采烈。祭祀以后曹文伯

(下图)

图 6 两译本“老百姓”一词的搭配(上图为张译本,下图为郁译本)

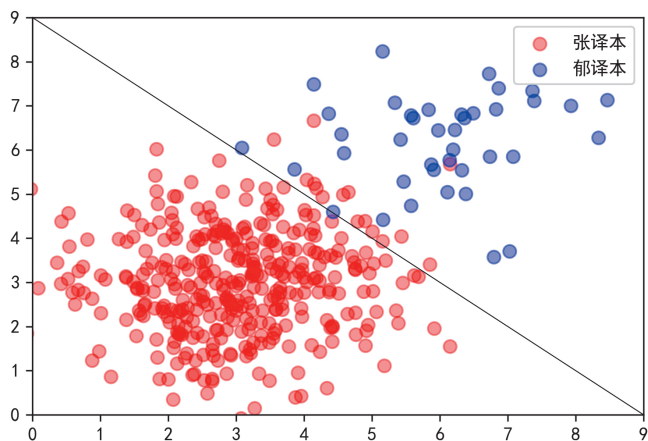


图 7 “老百姓”一词两译本分布强度对比

#### 4.4 讨论和启示

结合两位译者对原文原词语数量的压缩和增加的翻译行为,即从信息贡献度角度来分析同一词语如“大少爷”、“狐狸精”、“正人君子”和“老百姓”等称谓的增减,可以看出译者个人心理因素、教育修养、社会地位、政治偏好、文化习惯等在翻译作品中的“再创造”活动的影响,其潜在的社会、文化、宗教等心理因

素无疑通过对男女称谓的表达得到了投射或折射,或多或少地重塑或改变了原著者在原创作品里对主人公主动赋予的爱恨情仇和无意中流露的家国情怀,以及政治、民族和社会阶层意识等,不同程度地修正了原作中曾平亚、牛同义两位大少爷和牛素云等儿媳的角色性格、脾气、身份、气质、外表等形象。随着译语语义贡献度的变化,作品的特定语言文化意象得到不同程度的渲染和投射,作者、译者、译本和读者共同完成了对翻译文学作品的再创造,成功实现了文化的传播、变异和接受。当然,这不是本文要关注的主题。

本文尝试在语料库方法、语料库语言学和语料库翻译学应用领域引入信息贡献度这一概念。技术的引入是否符合学科研究的要求,关键在于它是否可应用于学科内较为经典的研究领域,如文学翻译研究。本文成功地进行了尝试,其结果证明这一技术与文学翻译研究构成了一个有机的闭环——在搭配语境的衬托之下,对比文本主题词的信息贡献度的高低预示着词汇语义迁移的方向或路径。

## 5 结语

信息贡献度方法是一个结合了文本语义结构的定义或方法,它不再仅仅独立地看待文本中的各种词汇。因此,它与经典而又传统的语料库技术工具相比有着较为本质的区别,同时它也有别于时下利用大数据进行训练的词向量模型方法。本方法的应用须借助Python进行,过程加载了多个与语料库语言学相关的工具包,它们发挥着不同的作用,有机地构成了本文研究所涉算法的各个环节。

结果业已呈现,当然有待同行的检验。在研究中引入新技术,其立足点应该是对既有经典技术的补充,为研究展示一个可供探索的维度。将信息贡献度方法应用于翻译文学和文化研究也仅仅是一次小小的创新尝试,期待方家批评指正,也期待有更多语言智能技术方法应用于翻译文学研究。□

## 参 考 文 献

- [1] Hamilton, W. L., Leskovec, J. & D. Jurafsky. Diachronic word embeddings reveal statistical laws of semantic change [J]. *arXiv preprint arXiv: 1605.09096*, 2016.
- [2] Montemurro, M. A. & D. Zanette. Towards the quantification of the semantic information encoded in written language [J]. *Advances in Complex Systems*, 2010(2).
- [3] Scott, M. PC analysis of key words—and key key words [J]. *System*, 1997(2).
- [4] 管新潮, 熊秋平. 翻译管理流程的界定与优化 [J]. *工业工程与管理*, 2012(2).
- [5] 姜孟, 鄢德平. 语义迁移的六类证据及其解释 [J]. *重庆工商大学学报*, 2013(1).
- [6] 林璐, 王旭. 语料库驱动下的英语副词 AGAIN 语义演变历时研究 [J]. *外语与外语教学*, 2019(2).
- [7] 刘知远, 刘扬, 涂存超, 孙茂松. 词汇语义变化与社会变迁定量观测与分析 [J]. *语言战略研究*, 2016(6).
- [8] 卢植, 胡健. 大数据时代背景下的现代语料库研制概览 [J]. *外语与翻译*, 2018(4).
- [9] 陆晓蕾, 王凡柯. 计算语言学中的重要术语——词向量 [J]. *中国科技词语*, 2020(4). (待发)
- [10] 秦洪武, 李婵, 王玉. 基于语料库的汉语翻译语言研究十年回顾 [J]. *解放军外国语学院学报*, 2014(1).
- [11] 王庆, 潘雪瑶. 英语同译词心理表征发展中的语义迁移机制 [J]. *当代外语研究*, 2018(3).
- [12] 王卫强. 认知视角下的跨文化语义迁移与偏移 [J]. *语文学刊*, 2010(7).
- [13] 王勇, 王李福, 邹辉, 何养明. 结合类别与语义贡献度的特征权重计算方法 [J]. *计算机工程与设计*, 2018(6).
- [14] 谢天振. 创造性叛逆: 争论、实质与意义 [J]. *中国比较文学*, 2012(2).
- [15] 许家金. 美国语料库语言学百年 [J]. *外语研究*, 2019(4).
- [16] 张伟华, 曾利沙. 论语境参数视阈下范畴语义转移的认知机制——兼论译者主客观互动的理据性 [J]. *外国语文*, 2014(6).

基金项目: 本文为上海市哲学社会科学规划项目一般课题“语言文化意象的翻译与特征建模研究”(项目编号: 2018BYY008)的阶段性成果。

收稿日期: 2019-09

通讯地址: 200083 上海市 上海外国语大学 (胡)

200240 上海市 上海交通大学外国语学院 (管)

## Revisiting Semantic Transfer in Literary Translation Studies: Using Key Keywords to Measure Informativeness

HU Jia-sheng<sup>1</sup> & GUAN Xin-chao<sup>2</sup>

( 1. Shanghai International Studies University , Shanghai 200083 , China; 2. School of Foreign Languages , STJU , Shanghai 200240 , China)

**Abstract:** This paper attempts to propose a new concept of informativeness as a key means to measure the contribution of lexis to a certain section or overall theme of translated works. This concept draws on inspirations from multidisciplinary knowledge as wide as information theory , corpus linguistics , translated literature , cultural communication , and cultural research of translation , etc. and bears significant theoretical and empirical values.

The key to traditional corpus research methods is the retrieval and identification of target words , while the key to big data research methods is the retrieval of big data and the training of word vector models. The former completes semantic construction in the subsequent process of interpreting the target word , while the latter constructs the semantic relationship through training the word vector model and obtain information for semantic interpretation in the training process.

In comparison with the traditional corpus methods and big data methods , informativeness integrates the affordances of these two approaches while avoiding the limitations regarding the huge amount of data. This is because informativeness-informed studies interpret semantic transfers in line with semantic structures of those lexis in texts , and thus have the potential to be applied in various semantic studies.

Using established algorithm available in Python

packages , this research sets out to compare cultural images constructed in different translations of literary works. Specifically , this research , based on two translations of Lin Yutang ' s English novel *Moment in Peking* ( aka *Jinghua Yanyun* in Chinese) , scrutinizes such terms as “Young Master” ( *dashaoye* ) , “She Fox” ( *hulijing* ) , “Gentleman” ( *zhengren junzi* ) and “Ordinary People” ( *laobaixing* ) from the perspective of informativeness , providing positive evidences to augment applicability of language intelligence in translation studies.

Our research showcases a closed loop from language intelligence and literary translation , meaning that comparing informativeness of key keywords can well indicate the directionality and paths of lexical semantic transfer.

We argue that informativeness-informed studies on semantic transfer depart from semantic structures of texts , retain the basic tenets of traditional corpus linguistic techniques , but chart new territories by combining Python-based big data methods with literary translation and corpus-based translation studies.

**Key words:** Semantic Transfer; Informativeness; Python; Literary Translation; Key Keywords; Cultural Image