

基于语料库的文化输出研究——以李子柒为例

杨瑞霞

(北京语言大学 高级翻译学院)

【摘要】近代以来，中国进入飞速发展。然而中国文化在对外传播的方面却不见起色。由于汉语的学习难度高，所以中国文化对外翻译几乎成为中国文化对外传播的主要方式。但是中国的文化负载词在目标语言中找到对应的名词，着实艰难。这也导致中国文化对外翻译的效果不甚理想。

然而在自媒体发展兴盛的年代，李子柒的横空出世却有力的传播了中国文化，许多中国美食得以通过李子柒的视频传播出去。更有甚者评价李子柒顶的上一百座孔子学院。

在传统文化对外翻译取得成效不甚显著的今天，是否应该探索一种新的文化外译方式？笔者希望以一种全新的思路看待李子柒的文化输出，供大家参考借鉴。并能够在这个全球互联的时代，以更好的方式来传播中国文化。

【关键词】语料库；文化外译；李子柒

研究背景

中国是典型的文化大国。自史前时期的炎黄二帝的传说开始，中华民族的概念就深深印刻在人们脑海中。中华文化也在这将近五千年的历史中不断得以发展与延续。然而自近代以来，中国文化一直不被西方世界所理解，人们对于中国的印象还停留在一个世纪以前那个落后的、贫穷的东亚病夫形象。所以如何构建新的中国形象？如何让中国文化再次走出去？

中国文化的走出去主要是以文字的形式，在这几千年的沉淀之中，中华文明早就铭刻在这卷帙浩繁的象形文字之中，所以中国文化对外翻译便成为了中国文化输出的重要途径。然而近年来随着科技信息的发展，超链接的诞生，人们获取信息的方式不再是单纯的文字形式，而是以自媒体的形式。特别是近些年来短平快的视频火爆全球，更有抖音（Tictok）这种主要以短视频为主的娱乐型app的兴盛。然而部分学者认为这种娱乐并

不是一种文化，更有学者呼吁相关政策能够节制人们的娱乐行为。然而随着李子柒等网红的视频在国外的火爆，我们不得不反思，难道这不是一种新形势下的文化对外翻译吗？

本文尝试回答三个问题：

1. 李子柒的视频能否代表中国文化？
2. 李子柒是否较好的传播了中国文化？
3. 李子柒的行为到底算不算一种文化对外翻译行为？

研究方法

以往的文化输出研究往往都是内省的，偏向于文科的研究方法，基于作者长期以来的实践以及主观意愿。笔者尝试使用语料库，一种由大数据支撑的方法对于该种翻译现象进行研究。

1. 李子柒的视频能否代表中国文化？

首先分析李子柒的视频内容。

李子柒的视频大多数以介绍中国美食为主。且李子柒的视频没有配音，没有字幕，只有

背景乐。视频中仅有的文字也只是菜名。所以直接通过视频内容来找到其中所隐含的文化内容是不实际的。因此，笔者尝试从与视频相关的文字信息找到视频隐含的文化内容。

通过相关查找，笔者发现视频下多有评论。评论是对于某一信息发表的看法，且评论多为当下看完之后的直观感受，所以评论可以反映视频所包含的内容。

李子柒其在国内等众多的视频网站中均有上线。在分析对比了一系列视频平台之后，笔者最终选定了微博这一平台。

微博是基于用户间简短实时信息共享的广播式社交媒体，尽管在 2016 年之后取消了字数限制，但博文仍然为短文本形式，其评论更是一大亮点，从评论中可以反映出大众对于某一信息的态度。

截至 2019 年 12 月 30 日，李子柒微博粉丝数量 21772984，共发表微博 388 篇。

在对其微博进行分析，并且充分考虑到本研究的体量之后，笔者选取其 2019 年发表的 33 篇视频微博，分别对其前 50 页的评论进行了抽取，得到共计 218025 字评论。

微博的受众主要为中文母语者，且覆盖各种年龄段的人群，其使用人群分布范围较为广泛。

抽取语料之后，采用 jieba 分词系统结合停用词表对文本进行分析处理，包括过滤掉一些只有在微博语言体系中才会使用的词语，例如：“微博抽奖”、“铁粉”、“转发”、“博主”等。之后利用词性标注系统过滤掉一些毫无意义的虚词、介词之后，最终得到了分词后词表集合。

得到的分词词表包含形容词、名词以及动词等。因为名词是最能够赋予意义的词语，动词形容词等在英文中存在很多对应的翻译，所以在不能够显著体现出中国文化的特征，所以在此分析名词。

指标描述：

笔者采用谷尼舆情图悦 picdata.cn 热词分析工具对数据进行分析。

以下为所选两个指标：

TF 热词指标：“TF 指标是一个词在文章中出现次数，出现的次数越多一般越重要，输出的词频信息只是参考，目前采用的分词方法是大词优先，不是以小词优先的，比如“改革”和“改革开放”都是词，在统计词频时，“改革开放”中的“改革”不会计入“改革”的词频，算两个不同的词。”

Score 热词权重指标：“Score 指标是指一个词在文章中重要性，主要由 TF 热词词频，IDF 倒转文档频率，other 其它三个指标决定。”

数据分析

对于名词进行分析之后，发现名词所代表的意义主要有以下三类：

第一类：表示人物的名词

称谓名词		
词语	热词词频 TF	Score 指标
李子柒	447	0.9337
奶奶	312	0.9224
姐姐	269	0.8962
仙女	183	0.889
婆婆	179	0.8846
柒姐	212	0.8691
神仙	143	0.8525
柒柒	135	0.8301
小姐	141	0.828
狗狗	94	0.8226
小狗	72	0.795
奇女子	40	0.7913
姑娘	72	0.782
女孩	65	0.7614
狗子	60	0.7604
外婆	41	0.7547

妈妈	56	0.7483
爷爷	40	0.7368
女神	34	0.7277
女孩子	35	0.7182
女子	37	0.7173
女侠	21	0.7107

1. 李子柒：在这类词语集合中频率最高的便是李子柒这个名字。还有一些比较家常的称呼，例如：“姐姐”、“小姐”、“姑娘”等。除了上述家常称呼之外还有“奇女子”、“神仙”、“仙女”、“女侠”这样颇具中国特色的词语。

2. 李子柒的奶奶：除了最常见的“奶奶”这样的称呼之外，还存在“婆婆”“外婆”这样的称呼。婆婆外婆这样的称呼在南方较为常见，可见李子柒视频的受众是较为广泛的，覆盖了中国的大部分人群。

3. 动物的出现：除了代表人物的名词之外，还存在一系列与狗相关的名词。在李子柒的视频中经常有小狗出境，所以评论中出现了与狗相关的词语。

4. 除了视频中出现的具体的人物动物形象之外，还出现了视频中没有对应形象的人物，比如“爷爷”这一词语。在依次浏览了原文本中“爷爷”出现的句段之后，笔者发现“爷爷”这一词出现的情况主要有以下两种

- 1) 和奶奶一起出现构成爷爷奶奶
- 2) 代指李子柒，描述李子柒能够做一些只有男性才做的粗活。

可见李子柒视频中体现出来的技能不仅仅只是美食，也包括一些中国古代男性才能做到的技能。从中也可以体现出李子柒的视频在内容上基本可以翻译中国古代农耕文化。

第二类：实体名词

实体名词		
词语	热词词频 TF	Score 指标
口水	123	0.8346
玉米	119	0.8272

藕粉	50	0.808
南瓜	70	0.8005
腊肉	62	0.7925
柚子	52	0.7918
火锅	70	0.7909
萝卜	68	0.7901
香菇	70	0.7875
青梅	56	0.7871
板栗	51	0.7845
水果	80	0.7836
地瓜	47	0.7791
粽子	46	0.7785
宝藏	49	0.7689
味道	66	0.7653
手套	52	0.7635
福利	59	0.7614
衣服	65	0.76
黄桃	57	0.7543
蛋黄	41	0.7538
辛夷花	55	0.753
番茄	46	0.7496
果子	38	0.7486
蜀绣	53	0.7475
镜头	50	0.7457
桃花	53	0.7439
蔬菜	50	0.7439
桃胶	56	0.7414
音乐	57	0.7406
辛夷	60	0.7387
草莓	40	0.7366
桃树	31	0.7357
谷子	29	0.7351
蘑菇	35	0.7317
辣椒	35	0.7295
榨菜	38	0.7291
刺绣	31	0.7275
身体	47	0.7257

豆浆	28	0.7214
栗子	27	0.7213
梅子	29	0.7184
鸭子	28	0.7176
稻谷	26	0.716
食材	21	0.7124
荷花	31	0.7121

经过分析发现实体名词中多为李子柒视频中出现的食物原材料，这些食物材料包括一些中国特有的食物原材料，包括藕粉、腊肉、火锅、板栗、粽子、辛夷花、桃胶、刺绣等能够体现中国特色、蕴含中国文化的食物。

第三类：时间和地点名词

时间名词		
词语	热词词频 TF	Score 指标
冬至	44	0.771
八月	48	0.7475
日子	56	0.7469
童年	47	0.7469
人间	48	0.7429
夏天	41	0.7281
冬天	56	0.756
大半夜	31	0.7554

地点名词		
词语	热词词频 TF	Score 指标
农村	127	0.8133
农活	61	0.8081
四川	111	0.807
院子	71	0.7942
仙境	66	0.7901
世外桃源	40	0.7519
家里	56	0.7495
山上	46	0.7471
山里	37	0.7429
老家	40	0.7388

绵阳	35	0.7345
----	----	--------

中国讲究二十四节气，所以时间也具有中国文化的特征。评论中我们可以看到诸如冬至这样的节气词语。

此外还有非常具有中国文化特征的地点名词。比如说中国的地名：四川、绵阳、仙境和世外桃源等。世外桃源语出陶渊明的《桃花源记》，这样的词语以前只有懂中文、且对中国古文有一定学习、了解陶渊明的所代表的那种田园文化的人才能想象出来，然而现在我们可以从李子柒的视频中感受到“世外桃源”这个词语背后的含义。从中可以看出，李子柒的视频确实承载着中国文化。

第四类：抽象名词

抽象名词		
词语	热词词频 TF	Score 指标
文化	88	0.7717
田园	55	0.7757
回忆	41	0.7226
诗意	30	0.7189
传统	41	0.7121

抽象名字往往是意义的集中体现，我们注意到在抽象名词这一栏中，文化的频数是最高的，“田园”一词位居第二，“诗意”位列第四。这三个词语都是非常具有文化气息的，这也和前面地点名词中提到的“世外桃源”相契合。此外还注意到“传统”一词也有着较高的词频，可见李子柒的视频确实蕴含着某种传统文化。除了上述词语外，还有“回忆”一词，和人物名词结合分析之后，笔者发现回忆一词和前文中人物名词中“奶奶”、“爷爷”、“外婆”、“婆婆”等联系较为紧密。这也从另一个侧面反映出李子柒的视频能够代表大多数人的农村生活，因而勾起了大众对于农村生活的回忆。

综上所述，在对各种能够较好的体现意义的名词进行分析后，发现李子柒的视频可以较好的体现中国人的文化内涵。

2. 李子柒是否较好的传播了中国文化？

文化输出除了需要有文化输出这一行为以外，我们也要看其文化输出的结果。一个没有影响力的文化输出不是一种好的文化输出。

研究方法

为了调查李子柒的视频在国外的影响程度，笔者对李子柒国外 Youtube 账号数据进行分析，并与部分中国官方账号进行对比，试图从对比中体现出李子柒的视频在国外影响力的大小。

数据分析

频道名称	订阅者数
李子柒	799 万
CGTN	104 万
CCTV 中国中央电视台	75.1 万

（数据截止截至 2019 年 12 月 30 日）

分析数据可以得出李子柒的频道订阅者人数远大于 CGTN 和 CCTV 中国中央电视台这样的中国主流对外宣传媒。

而李子柒在 Facebook 的粉丝数 296 万，获赞总数为 203 万。从李子柒在另外一个国外主流媒体 Facebook 上的数据也可以侧面反映出李子柒在国外的受欢迎程度。

根据以上数据的纵向以及横向对比，我们可以看到李子柒在海外的影响力相对而言非常大，其对于中华文化的宣传效果比中国中央电视台的主流对外宣传媒介要好。

3. 李子柒视频的传播到底算不算中国文化外译？中国文化外译是对于中国传统文化对外翻译的简称。关于李子柒的视频能否代表中国传统文化，在第一部分已经进行过详细叙述。“对外”二字，在第二个问题中也已经详细叙述。最后回归到“翻译”这个问题上。

首先来看一下翻译的定义：

我们采用巴尔胡达罗夫的说法，巴尔胡达罗夫认为：“翻译是一种语言的言语产物（话语）在保持内容，即意义不变的情况下改变为另外一种语言的言语产物的过程。”（蔡毅 1995(06):9-12.）

上述观点狭义的集中在语言上，笔者尝试从雅各布森的从言语符号的意义上提出的三种译域观点出发，对李子柒的这视频进行分析。

罗曼·雅各布森（Roman Jakobson, 1892—1982）从符号学的观点出发，看翻译的三大类别，即语内翻译（intralingual），语际翻译（interlingual translation）和符际翻译（intersemiotic translation）。语内翻译是指用一个语言中用一些语言符号解释另一些语言符号，就是人们所说的“改变说法”，语际翻译是指两种语言之间的解释，即用另一种语言的语符来解释一种语言的语符。符际翻译，就是通过非语言的符号系统解释语言符号，或用语言符号解释非语言符号（李静 2009(01):138-140）。

雅各布森对于翻译的看法为翻译定义提供了一种全新的视角。

他认为：“准确的翻译取决于信息的对等。在语内翻译中，是用一个语符单位代替另一个语符单位；对一个词的翻译可以选用同义词，也可以采用迂回表示法。但一般来说，用一次不可能是完全对等的词。在语际翻译中符号于符号之间一般也没有完全的对等关系。人们不是用一种语言的语符来代替另一种语言中单个的符号，而是代替“更大的单位”，即雅各布森所称的“信息”（message），是用信息代替信息。”（李静 2009(01):138-140）

这里雅各布森所说的“信息”究竟指代的是什么？在下文的陈述中我们可以得知雅各布森对于信息的解释依然是停留在语言符号上的，因为雅各布森受限于当时的环境。

当时人们之间的信息交流主要以语言和文字为主，所以研究学者们研究的问题也是以翻译文字为主，也就说所说的符号。然而从当代的角度看待翻译问题，信息的传递方式变得多样化，多媒体文本对于信息的传输方式有非常大的影响。一张图片、一段视频都能够传递某个信息。雅各布森所说的信息在新时代下应该赋予不一样的意义，符号是什么？如果符号能够传递信息的话，那么图片便符合这样的定义，包括当代年轻人之间流行的表情包文化，表情包也是能够传递信息的符号。所以从某种意义上来说，视频信息也算是一种符号，李子柒将某种信息翻译成了视频符号，这种符号能够更加广泛的被别人接受，所以才有了李子柒的视频在海外的广泛传播。

那接下来讨论李子柒翻译了什么信息。在第一部分对于李子柒的视频内容的讨论中，我们可以从评论的文字中看出李子柒的视频主要是对于中国传统文化的一种再现。包括如何制酱油，如何做家具……笔者认为李子柒将原来不可感知，仅仅存在于文字或者是人们口口相传的语言这种不可感知的过程，翻译成为了视频；李子柒还将一些用言语难以描述的信息，翻译成了视频。最典型的一个例子便是对于“世外桃源”以及“仙女”的演绎。以前，人们对于世外桃源的想象仅仅停留在陶渊明的诗词之中，是“种豆南山下，草盛豆苗稀”这样文绉绉的诗句。然而经过李子柒的翻译，我们可以更加深刻的感受到什么是“世外桃源”，什么是“田园生活”。即使是不懂得中文的非中文母语者，也可以感受到其中陶渊明笔下的“世外桃源”。面对这样的一种视频信息，文化传播的力度自然是大于晦涩难懂的纯文本。

虽然有的学者认为现代这种多样化的信息传递方式导致翻译的边界变得越来越模糊，这是对于翻译的一种不尊重。

但是笔者以为，翻译的目的就是为了更好的交流，以一种更加容易被大众接受的方式来进行翻译，会起到事半功倍的效果。在如今这个全球化的时代，信息的传播速度飞快，如果我们能够端正自己的态度，积极学习采纳这样的建议，中国文化在海外的传播也会有不一样的图景。

结论

李子柒能够代表中国传统文化，且其视频在海外确实起到了传播中国文化的作用，而她的视频从雅各布森符际翻译的观点出发，也可以被理解成一种文化对外翻译行为。

对于大部分热衷于中国文化对外输出的译者来说，他们既了解中国文化，同时对于外国文化也有较好的理解，能够理解二者之间的文化差异。在中国文化对外输出的过程，这些译者在文字驾驭方面和非文字工作者相比，有着较强的能力，但是在提升文字水平的同时也应该考虑采用一些多媒体手段。毕竟文字能够传递的信息是有限的，在这个通过富文本（视频、图像、声音……）传递信息已经成为主流的时代，译者也应该考虑自己传递信息的方式，以一种更加流行，大众更加接受的方式来传递信息，更好的输出中国文化。

参考文献：

- [1] 蔡毅. 关于国外翻译理论的三大核心概念——翻译的实质、可译性和等值[J]. 中国翻译, 1995(06): 9-12.
- [2] 李静. 雅各布森翻译理论研究[J]. 文学教育(上), 2009(01): 138-140.