

# 纪录片的多模态话语分析

## ——以《走进大凉山》为例

沈思, 穆红

大连理工大学外国语学院, 辽宁 大连

收稿日期: 2022年12月19日; 录用日期: 2023年2月9日; 发布日期: 2023年2月20日

### 摘要

本文通过竹内亮导演制作的《走近大凉山》为例展开多模态纪录片分析, 通过ELAN标注视频中的8种模态(画面配置、画像、空间距离、身体动作、手部动作、头部动作、语言、音乐), 共发现以下2条规律: 1) 《走近大凉山》的各种模态中语言层占比最高, 其次为身体动作层, 第三者为空间距离层。2) 在语言层中主要使用C(汉语)、J(日语)等来表示; 身体动作层BGZ(坐)、BGZL(站)等来表示; 手部动作HMZ(中动作)、HMS(向上动作)等来表示; 头部动作TMZT(转身)、TMTT(抬头)等来表示; 空间距离层ZN(近距离)、ZF(远距离)等来表示; 音乐层SM(主观音乐)、OM(客观音乐)等来表示; 图像层IT(文字)、ITV(文字和视频)等来表示; 画面配置AP(人)、APS(人与景观)等来表示。

### 关键词

多模态话语, 纪录片, ELAN

# Multimodal Discourse Analysis of Documentary

## —Taking “Walking into Daliang Mountain” as an Example

Si Shen, Hong Mu

College of foreign languages, Dalian University of Technology, Dalian Liaoning

Received: Dec. 19<sup>th</sup>, 2022; accepted: Feb. 9<sup>th</sup>, 2023; published: Feb. 20<sup>th</sup>, 2023

### Abstract

In this paper, the multi-modal documentary analysis is carried out by taking “Approaching Liangshan” produced by director Takeuchi Ryo as an example, Eight modes (picture configuration, por-

trait, space distance, body movement, hand movement, head movement, language and music) in the video are marked through ELAN. The following two laws are found: 1) Among the modes of “Approaching Liangshan”, the language layer accounts for the highest proportion, followed by the body movement layer, and the third is the space distance layer. 2) In the language layer, C (Chinese) and J (Japanese) are mainly used to express; body movement layer BGZ (sit), BGZL (stand), etc. Hand movements HMZ (middle movement), HMS (upward movement), etc.; Head movement TMZT (turn around), TMTT (head up), etc. Spatial distance layer ZN (close distance), ZF (long distance), etc. Music layer SM (subjective music), OM (objective music), etc. Image layer IT (text), ITV (text and video) and so on; the picture is represented by AP (People) and APS (People and Landscape).

## Keywords

Multimodal Discourse, Documentary, ELAN

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

中国是世界上人口最多的发展中国家。由于出口主导的增长战略,沿海地区实现了快速现代化,但内陆中西部地区的经济发展相对缓慢,地区间出现了巨大的经济差距。因此,国内依然存在大量贫困人口。中国政府多年来以减贫为目标,实施了各种脱贫对策和政策,脱贫是我国政府重要国策之一。其中位于我国四川省的大凉山地区,因交通不便、发展缓慢等原因,是公认的中国最贫困的地区之一。大凉山地区居住着以彝族为首的众多少数民族。日本纪录片导演竹内亮以大凉山的情况为背景制作了纪录片《走进大凉山》,该片在世界最大的视频平台 YouTube 上的播放次数为 3,555,414 万次,评论数为 15,144 件(截止至 2022 年 12 月 19 日)。这部纪录片备受瞩目,获得了观众的一致好评。

对于纪录片的分析,以往的研究多从宣传学、美学、叙事策略等分析视角展开了大量的研究,本研究从多模态领域,剖析该纪录片如何记录大凉山的情况,探讨如何运用各种模态呈现想要传达的理念和信息。

通过纪录片等媒体对中国精神进行传播,构建中国国家形象是至关重要的。考察外国导演如何传达中国的情况,分析该纪录片的表现方法,可以为通过纪录片传达中国国家形象提供参与与启示。

## 2. 文献综述

纪录片的基本特征可以说是真实地记录世界。纪录片指的是除剧情片以外的所有电影。纪录片本质上是创作者对眼前发生的事情的观察和记录的成果(任远 2002) [1]。纪录片作为跨文化宣传的重要载体具有塑造国家形象、宣传民族文化、提升国家文化软实力的责任和使命(武新宏·吴保平 2021) [2]。

关于纪录片的研究虽然开展了很多,但其中大多数都是从传达学和叙事策略的角度进行的分析和纪录片本身的内容分析。何建平,赵毅岗(2007) [3]对中西纪录片的“文化折扣”(a cultural discount)现象进行了深入研究,并阐述了中国纪录片为开拓欧美市场如何减少“文化折扣”的建议。对于纪录片起到传达国家形象的作用这一观点,许多学者达成了一致。沈悦,尹如歌(2018) [4]以一带一路为背景,论述了纪录片国家形象的构建及跨文化传播的对策。

总的来看,关于纪录片的研究中媒体学方面的研究较多,从语言学的视角分析纪录片的研究较少,而

从多模态分析的角度分析纪录片的研究则更少。代表的研究主要有 Yao, Yinyan-Zhuo, Yanfen (2018) [5], Z Ullah (2019) [6], 刘煜, 张红军(2018) [7]等研究。Yao Yinyan, Zhuo Yanfen (2018) [5]对中国杭州市宣传视频进行多模态分析, 根据视觉语法及符号间互补性框架, 分析各种符号学资源如何解释、如何协同的协同作用。Z Ullah (2019) [6]以纪录片《A Girl in the River》为分析对象, 结合性别研究、女权主义、电影语言、电视电影功能的多模态视差分析, 探讨了视觉、听觉、动作、编辑和其他颜色模态的存在情况。

另一方面, 在中国也有以纪录片为对象进行多模态分析的研究。刘煜、张红军(2018) [7]从多模态分析的视角, 探讨了政论纪录片如何感知、认知国家形象, 进而推动认同。通过多模态的四重实践, 即表现给受众感觉的提示实践、致力于意义传达的话语实践、扩展时空维度的传达实践、形成深刻共鸣的社会实践, 发挥纪录片的最大宣传作用。

随着科技的发展, 特别是多媒体技术的发展, 语音、影像、视频等符号学资源的综合应用, 可以说信息更加生动、更加具体。因此, 近年来传统话语分析逐渐从语言向多模态话语分析变化(朱永生 2007) [8]。Kress 和 Van Leeuwen (1996) [9]认为, 所谓多模态是指通过包括语言、手势、影像、符号、文本等媒体的多个模式、形式来表现意义。Lemke (2002) [10]指出, 语言、手势、音乐等符号资源系统可以相互协调地定义为多模态。此外, Van Leeuwen (2005) [11]指出, 多模态是指语言、影像、声音等不同的符号模式在艺术创作和交流过程中的相互配合。Baldry 和 Thibault (2006) [12]认为, 多模态描述了多个不同的符号资源系统, 这些资源系统被统一地布置并统一地上下文化, 以帮助形成特定的语义。

### 3. 研究方法

#### 3.1. 研究对象

本研究的分析对象《走进大凉山》是由解读中国工作室和纪录片导演竹内亮共同制作的, 2021年4月28日起在中国和日本的各电视平台和视频平台上播出。到2022年12月19日为止, 播放次数为3,555,414次。竹内亮曾在NHK等日本媒体工作多年, 后移居江苏省南京市。此后, 他对中国进行了更深入的观察和记录, 制作了《我住在这里的理由》、《好久不见, 武汉》等纪录片纪行节目, 其独特的风格在中国和日本博得好评。在《走近大凉山》中并没有没有消极地描写贫困生活, 而是以日本人一边旅行一边自然地融入当地人的生活视角来展开描写。《走近大凉山》在日本媒体上发布, 这是很多日本人了解中国脱贫故事的机会。

#### 3.2. 分析方法

本研究依据 Norris 多模态相互分析理论对纪录片《走进大凉山》进行分析。Norris (2004) [13]考察了在交流场合人们如何使用交流模式进行社交, 并将使用的模式分为“语言”、“空间距离”、“身体动作”、“手部动作”、“头部动作”、“眼神”、“音乐”、“印刷品”、“画面配置”9大类。其中, 1) 语言: 指参加交流场景的人的对话。作为使用频率最高、最重要的交流模式被广泛使用。2) 空间距离: 指分析场景中记者与被采访者的关系、被采访者之间的关系、人与环境的关系等空间距离。3) 身体动作: 指参加交流场景的人的动作等。主要分为身体位置和朝向的方向。4) 手部动作: 指手的运动轨迹以及手部动作的身体语言。5) 头部动作: 指交流场景中进入者头部的移动。主要分为方向变化、有节奏的点头、姿势移动、眼神移动四种。6) 眼神: 作为信息的重要传递方法, 分为观察内容、观察方向、观察强度。7) 音乐: 作为听觉模式之一被广泛使用。有在编辑动画时插入的歌曲, 也有为了表达自己的感想而自然演唱的歌曲。8) 印刷品: 指媒体资料、图表等纸面化的文本。9) 画面配置: 指的是画面的布局, 也包括画面上出现的人、物、景观的布局。本研究基于 Norris 多模态分析理论, 从语言层、身体动作层、手部动作层、头部动作层、音乐、图像、空间距离、画面配置等方面对纪录片的视频进行多模态分析。

### 3.3. 分析步骤

在多模态分析中, 转录是必不可少的, 即以文本的形式按照提示的顺序整理并记录动态图像的构成要素或基本信息。Baldry Thibault (2006) [12]提出了动态图像分析和转录两种方法。即宏观分析和微观分析。宏观分析的转录在于获得视差的基本结构。可以通过构成置换的单元之间的关系来解释置换的语义生成过程。单元主要包括阶段、集群和帧。微观分析的转录是对语义生成过程中使用的符号资源进行详细的描述, 但转录一定要有选择性, 只转录与分析有关的模式, 避免转录的无限详细描述和漫无目的性。

本研究主要对以上 Baldry Thibault (2006) [12]提出的转录和分析方法进行动态图像置换, 同时也参考其他相关学者的研究结果。例如, 在分析层次中, 为了更清楚层次, 综合 Iedema (2001) [14]和 Baldry Thibault (2006) [12]的成果, 将视频分为阶段、类型步骤、场景、帧这四个层次。阶段是指一系列共同作用的符号模式, 在话语批评分析的特定时间内保持整合的一致性, 一个阶段所表达的意义区别于该作品中其他阶段所表达的意义。类型步骤是指阶段的开头、中间、末尾的时间点。场景是指用两个或多个镜头来表现一个地点、一段时间内发生的一系列事件。帧是指透镜内的一张静态图像。

本研究采用 ELAN 软件进行视频的注释和转印, 分析视频各模式的使用情况。ELAN 可以向音频或视频记录任意添加文本注释。注释可以是语句、单词、翻译或媒体上观察到的任何特征的描述。

## 4. 多模态呈现情况分析

本研究首先将视频导入 ELAN 软件, 根据 Norris (2004) [13]的分类对各场景进行注释, 分析各模式的呈现情况。张德禄(2009) [15]认为视频多模态应当分为语言、身体、非身体三种。语言是指伴随着语言本身和语言。非语言包括身体和非身体。**错误! 未定义书签。**本研究按照语言、身体、非身体 3 个分类, 对视频的各场景进行模式注释并进行分析。计算出动画的各模式的时间及比例, 如表 1 所示。其中, 语言注释层的注释时间总计为 2673.064 秒, 占注释总时间的 85.72%。身体注释层合计为 2340.334 秒, 占注释总时间的 75.05%。身体注释层合计为 2340.334 秒, 占注释总时间的 75.05%。换句话说, 语言可以说是本视频中时间最长、信息量最高的模式。表 2 为每一部分详细分类。

**Table 1.** Usage and proportion of each mode

**表 1.** 各模态使用情况和比例

注释层	总时间(单位秒)	总注释时间(单位秒)	时间百分比(单位%)
语言	2673.064		85.72%
身体	2340.334	2968	75.05%
非身体	1638.84		52.55%

**Table 2.** Specific usage and proportion of each mode

**表 2.** 各模态的具体使用情况和比例

注释层	总时间	时间百分比
语言层	2673.064	85.72%
身体动作层	2024.696	64.92%
手部动作层	219.051	7.02%
头部动作层	96.587	3.10%
空间距离层	756.457	24.26%
音乐层	688.246	22.07%
图像层	194.137	6.23%

基于 Norris 的分类角度, 从“语言”、“空间距离”、“身体动作”、“手部动作”、“头部动作”、“眼神”、“音乐”、“印刷品”、“画面配置”9类对视频的模式进行分类, 各个模式的具体使用情况和时间比如表 2 所示。

#### 4.1. 语言层

在语言层, 纪录片《走近大凉山》中多个语言单独或同时出现的情况较多。视频中出现的语言主要有汉语、日语、彝族语、英语四种。各语言的对应关系如表 3 所示。具体来看, 该纪录片中单独使用的语言是汉语、日语、英语、彝族语。汉语(C)的使用时间合计为 1444.63 秒, 占总时间的 46.32%。日语(J)的使用时间合计为 731.067 秒, 占全体时间的 23.44%。英语(E)的使用时间合计为 5.001 秒, 占总时间的 0.16%。彝族语(Y)的使用时间合计为 4.884 秒, 占全体时间的 0.16%。由此可见, 汉语是使用频率最高的语言。

**Table 3.** Usage time and proportion of languages in the language layer

**表 3.** 语言层中各语言的使用时间和比例

具体分类	总时间(单位秒)	时间百分比(单位%)
C (汉语)	1444.63	46.32%
J (日语)	731.067	23.44%
CJ (汉日)	361.744	11.60%
CJY (汉日彝)	42.595	1.37%
CY (汉彝)	53.381	1.72%
CE (汉英)	29.762	0.95%
E (英语)	5.001	0.16%
Y (彝族语)	4.884	0.16%

视频中可以看出, 竹内亮在与村里人交流或一个人说话时, 使用的是汉语。例如, 下面的图 1 是竹内亮和村里的人用汉语说话的场景。其次, 纪录片中使用较多的语言是日语。主要是在竹内亮的旁白或者和别人说话的场合中使用日语。下面的图 2 是竹内亮用日语说明和说话的场景。



**Figure 1.** Scenario of using Chinese alone

**图 1.** 单独使用汉语的场景



**Figure 2.** Scenario of using Japanese alone

**图 2.** 单独使用日语的场景

另外, 分析发现单独使用彝族语和英语的情况并不多。约占整体的 0.3%左右。主要是在村里人和外国来的足球教练说话的场合中, 使用的是彝族语和英语。图 3 是单独使用彝族语的场景, 图 4 是单独使用英语的场景。



Figure 3. Scenario of using Yi language alone  
图 3. 单独使用彝族语的场景



Figure 4. Scenario of using English alone  
图 4. 单独使用英语的场景

另一方面, 两种语言同时使用时, 主要是“汉语和日语”、“汉语和彝族语”、“汉语和英语”的组合, 分别为 361.744 秒、53.381 秒、29.762 秒, 其比例分别为 11.60%、1.72%、0.95%。“汉语和日语”中所说的场景, 主要是竹内亮用汉语和日语说话的场景。图 5 是竹内亮用汉语和日语说话的场景。用“汉语和彝族语”说话的场景主要是采访刺绣基地工作人员时的场景。图 6 是竹内亮与刺绣基地工作人员交谈的场景。



Figure 5. Scenario of using Chinese and Japanese  
图 5. 用汉语和日语说话的场景



Figure 6. Scenario of using Chinese and Yi language  
图 6. 用汉语和彝族语说话的场景

用“汉语和英语”说话的场景主要是采访来自外国的足球教练时的场景。图 7 是竹内亮与足球教练交谈的场景。另外, 两种以上的语言同时使用的是“汉语、日语、彝族语”。图 8 为竹内亮在彝族大妈家聊天时, 说汉语和彝族语时, 不经意间使用了日语。



Figure 7. Scenario of using Chinese and English  
图 7. 用汉语和英语说话的场景



Figure 8. Scenario of using Chinese, Japanese and Yi language  
图 8. 用汉语、日语、彝族语说话的场景

## 4.2. 身体动作层

身体动作是指交流场景中进入者身体的伸展和弯曲等动作。本研究将交流场合的动作分为“坐、走、站、蹲、跑、跳、躺”。其结果如表 4 所示。

Table 4. Proportion of different categories of body movements

表 4. 身体动作各分类的比例

具体分类	总时间	时间百分比
BGZ (坐)	807.351	25.89%
BGZL (站)	798.181	25.60%
BGXZ (走)	268.778	8.62%
BGXD (蹲)	83.009	2.66%
BGPB (跑)	47.7	1.53%
BGTW (跳)	18.994	0.61%
BGT (躺)	0.683	0.02%

首先, “坐(BGZ)”的注释时间合计为 807.351 秒, 占总时间的 25.89%。“站立(BGZL)”的注释时间合计为 798.181 秒, 占总时间的 25.5%。“走(BGXZ)”的注释时间合计为 268.778s, 占总时间的 8.62%。“蹲(BGXD)”的注释时间合计为 83.009 秒, 占总时间的 2.66%。“跑(BGPB)”的注释时间合计为 47.7 秒, 占总时间的 1.53%。跳(BGTW)的注解时间合计为 18.994 秒, 占总时间的 0.61%。“躺(BGT)”的注

释时间合计为 0.683 秒, 占总时间的 0.02%。总体来说, “坐”和“站”是使用频率最高的, 其次是“走”。下面的图 9 是身体动作的各场景。



Figure 9. Scenarios of body movements  
图 9. 身体动作的各场景

### 4.3. 手部动作层

手部动作是指手的运动轨迹以及手部动作所产生的身体语言。作为分析对象的纪录片很少有隐喻性的手势, 因此本研究将交流场景中的手势分为向上动作、中动作、向下动作。其结果如表 5 所示。

Table 5. Proportion of each classification of hand movements  
表 5. 手部动作各分类的比例

具体分类	总时间	时间百分比
HMZ (中部动作)	112.972	3.62%
HMS (向上动作)	88.298	2.83%
HMX (向下动作)	17.781	0.57%

具体来看, 上面的动作(HMS)的注释时间合计为 88.298 秒, 占总时间的 2.83%。中的动作(HMZ)的注释时间合计为 112.972 秒, 占总时间的 3.62%。发现向下动作(HMX)的注释时间合计为 17.781 秒, 占总时



间的 0.57%。即表示中的动作的使用频度最高, 其次是向上的动作。图 10、图 11、图 12 是向上的手部动作、中部手部动作、向下的手部动作的场景。



Figure 10. Scenario of upward hand action

图 10. 向上的手部动作



Figure 11. Scenario of middle hand action

图 11. 中部的上部动作



Figure 12. Scenario of downward hand action

图 12. 向下的手部动作

#### 4.4. 头部动作层

头部动作指的是进入交流场景的人头部的移动。本研究将头部动作分为“点头、转身、抬头”。其结果如表 6 所示。

首先, “点头(TMDT)”的注释时间合计为 13.81 秒, 占总时间的 0.44%。“回头(TMZT)”的注释时间合计为 60.539 秒, 占总时间的 1.94%。“抬头(TMTT)”的注释时间合计为 22.238 秒, 占总时间的 0.71%。换言之, “回头”是使用频率最高的, 其次是“抬头”。下面的图 13 是头部动作的各场景。

Table 6. Proportion of head movement categories

表 6. 头部动作各分类的比例

具体分类	总时间	时间百分比
TMZT (转身)	60.539	1.94%
TMTT (抬头)	22.238	0.71%
TMDT (点头)	13.81	0.44%

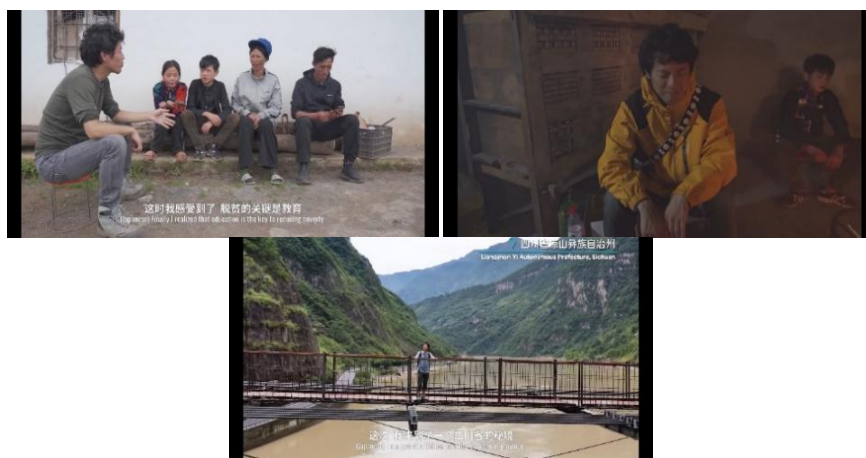


Figure 13. Scenarios of head movements

图 13. 头部动作的各场景

#### 4.5. 空间距离层

空间距离层是指分析场景中记者与被采访者的关系、被采访者之间的关系、人与环境的关系等空间距离。本研究将其分为“近距离、中距离、远距离”。其结果如表 7 所示。

Table 7. Proportion of each classification of spatial distance layer

表 7. 空间距离层各分类的比率

具体分类	总时间	时间百分比
ZN (近距离)	596	19.11%
ZF (远距离)	111.492	3.58%
ZM (中距离)	48.965	1.57%



Figure 14. Scenario of short distance, medium distance and long distance

图 14. 近距离、中距离、远距离的场景

首先,“近距离(ZN)”的注释时间合计为 596 秒,占总时间的 19.11%。“中距离(ZM)”的注释时间合计为 48.965 秒,占总时间的 1.57%。“远距离(ZF)”的注释时间合计为 111.492 秒,占总时间

的 3.58%。也就是说,“近距离”是使用频率最高的,其次是“远距离”。图 14 是近距离、中距离、远距离的场景。

#### 4.6. 音乐层

音乐层既有制作视频时插入的歌曲或歌曲(主观音乐),也有不是制作者插入的歌曲,而是在拍摄时处于该场景的音乐(客观音乐)。本研究将其分为“主观音乐、客观音乐”。其结果如表 8 所示。

**Table 8.** Proportion of various categories of music layer

**表 8.** 音乐层各分类的比率

具体分类	总时间	时间百分比
SM (主观音乐)	582.014	18.66%
OM (客观音乐)	106.232	3.41%

具体来看,“主观音乐(SM)”的注释时间合计为 582.014 秒,占总时间的 18.66%。“客观音乐(OM)”的注解时间合计为 106.232 秒,占总时间的 3.41%。也就是说,“主观音乐”是使用频率最高的音乐,其次是“客观音乐”。

#### 4.7. 图像层

图像层是用于补充的信息,如图表、标记、文字、动画等。本研究将其分为“文字、视频、图、标记”。其结果如表 9 所示。

**Table 9.** Proportion of each classification of image layer

**表 9.** 音乐层各分类的比率

具体分类	总时间	时间百分比
IT (文字)	93.731	3.01%
ITV (文字和视频)	37.064	1.19%
ITP (文字和图)	33.88	1.09%
IP (图)	25.415	0.82%
IM (标记)	4.047	0.13%

首先,“文字(IT)”的注释时间合计为 93.731 秒,占总时间的 3.01%。“文字和视频(ITV)”的注释时间合计为 37.064 秒,占总时间的 1.19%。“文字和图(ITP)”的注释时间合计为 33.88 秒,占总时间的 1.09%。“图(IP)”的注释时间合计为 25.415 秒,占总时间的 0.82%。“标记(IM)”的总注释时间为 4.07 秒,占总时间的 0.13%。总体而言,“文字”是使用频率最高的,其次是“文字和视频”。图 15 是用文字补充信息的场景。

#### 4.8. 画面配置

画面配置指的是画面的布局,也包括画面上出现的人、物、景观的布局。按其组合分为“人”、“物”、“景观”、“人与物”、“人与景观”、“人与物与景观”“景观与物”七类。其时间和比例如表 10 所示。总体而言,“人”的画面布局是使用频率最高的,其次是“人与景观”。画面配置的具体例如图 16 所示。



Figure 15. Text scene  
图 15. 文字场景

Table 10. Proportion of various categories of screen configuration  
表 10. 画面配置各分类的比例

具体分类	总时间	时间百分比
AP (人)	1580.018	50.67%
APS (人与景观)	443.36	14.22%
APT (人与物)	327.236	10.49%
APTS (人与物与景观)	26.593	0.85%
AS (景观)	297.442	9.54%
AT (物)	152.314	4.88%
ATS (景观与物)	24.265	0.78%



Figure 16. Scenario of screen configuration  
图 16. 画面配置的各场景

## 5. 结论

本文通过竹内亮导演制作的《走近大凉山》为例展开多模态纪录片分析, 通过分析视频中的 8 种模态之间的相互作用, 发现在该纪录片中模态之间的相互作用关系可以总结为以下 2 条规律。

1) 《走近大凉山》的各种模态中语言层占比最高, 其次为身体动作层, 第三者为空间距离层。

2) 在语言层中主要使用 C (汉语)、J (日语) 等来表示; 身体动作层 BGZ (坐)、BGZL (站) 等来表示; 手部动作 HMZ (中动作)、HMS (向上动作) 等来表示; 头部动作 TMZT (转身)、TMTT (抬头) 等来表示; 空间距离层 ZN (近距离)、ZF (远距离) 等来表示; 音乐层 SM (主观音乐)、OM (客观音乐) 等来表示; 图像层 IT (文字)、ITV (文字和视频) 等来表示; 画面配置 AP (人)、APS (人与景观) 等来表示。

因此, 在以上 2 条规律的辅助之下, 该介绍中国的记录篇在世界范围内获得了较高的评价。而我国可以通过借鉴以上关系, 更好地推动中国文化走向世界, 讲好中国故事。

## 参考文献

- [1] 任远. 非虚构是纪录片最后防线——评格里尔逊的“创造性处理”论[J]. 现代传播, 2002(6): 39-42.
- [2] 武新宏, 吴保平. 后疫情时代纪录片“他塑”中国形象——以日本导演竹内亮中国疫情题材纪录片为例[J]. 电影新作, 2021(4): 83-87.
- [3] 何建平, 赵毅岗. 中西方纪录片的“文化折扣”现象研究[J]. 现代传播(中国传媒大学学报), 2007(3): 100-104.
- [4] 沈悦, 尹如歌. 中国纪录片的国家形象建构与跨文化传播——“一带一路”视阈下的再思考[J]. 云南民族大学学报(哲学社会科学版), 2018, 35(2): 31-38. <https://doi.org/10.13727/j.cnki.53-1191/c.2018.02.005>
- [5] Yao, Y.Y. and Zhuo, Y.F. (2018) A Multimodal Discourse Analysis of the Promotional Video of Hangzhou. *English Language Teaching*, **11**, 121-131. <https://doi.org/10.5539/elt.v11n10p121>
- [6] Ullah, Z. (2019) Unfolding the Honour Killing Documentary “A Girl in the River”: A Multimodal Telecinematic Analysis. *Journal of Gender and Social Issues*, **18**, 83-96.
- [7] 刘煜, 张红军. 政论纪录片塑造国家形象的多模态话语分析[J]. 现代传播(中国传媒大学学报), 2018, 40(9): 118-122.
- [8] 朱永生. 多模态话语分析的理论基础与研究方法[J]. 外语学刊, 2007, 5(1): 82-86.
- [9] Kress, G. and Van Leeuwen, T. (1996) *Reading Images: The grammar of Visual Design*. Routledge, London.
- [10] Lemke, J.L. (2002) *Travels in Hyper-Modality*. Routledge, London. <https://doi.org/10.1177/147035720200100303>
- [11] Van Leeuwen, T. (2005) *Introducing Social Semiotics*. Routledge, London.
- [12] Baldry, A. and Thibault, P.J. (2006) *Multimodal Transcription and Text Analysis*. Equinox, London.
- [13] Norris, S. (2009) Modal Density and Model Configurations: Multimodal Actions. In: Jewitt, C., ed., *The Routledge handbook of multimodal Analysis*, Routledge, London, 78-90.
- [14] Iedema, R. (2001) Analysing Film and Television: A Social Semiotic Account of Hospital: An Unhealthy Business. In: van Leeuwen, T. and Jewitt, C., eds., *Handbook of Visual Analysis*, Sage, London.
- [15] 张德禄. 多模态话语分析综合理论框架探索[J]. 中国外语, 2009, 6(1): 24-30.