

A Discourse Analysis of the Voice-over Narration Power in Ethnic Minorities Documentaries: A Case Study of "Image Journal of Fifteen Endemic Minorities in Yunnan, China"

Yihe Zhao

*School of Theater, Film and Television, Communication University of China, Beijing, China
19960742906@163.com*

Abstract. In many countries, documentaries about ethnic minorities are produced by teams from the dominant ethnic group. Beyond the unconscious portrayal of positions during planning, filming, and editing, the choice of voice-over narration strategies can also invisibly reinforce unequal power relations. This paper examines the Chinese documentary *Image Journal of Fifteen Endemic Minorities in Yunnan, China*, and explores how voice-over narration influences power distribution through linguistic manipulation. Employing discourse analysis, this study examines the imbalance in discourse power that results from linguistic strategies, including the use of third-person pronouns, deontic modality, and abstract adjectives. Quantitative analysis further highlights the marginalization of ethnic minority voices, with the mother tongue appearing in less than 2% of the film's screen time. The results suggest that narration, as a discursive strategy, shapes power relations and suppresses the subjectivity of ethnic minorities. This study contributes to the connection between documentary research and linguistic analysis, advocating for more ethical and diverse narrative practices.

Keywords: ethnic minority documentary, voice-over narration, discourse analysis, discourse power, linguistic strategies

1. Introduction

Since the 1960s and 1970s, the rise of *cinéma vérité* and direct cinema has profoundly reshaped documentary narrative and production practices. The once-dominant, authoritative “voice of God” narrative gave way to more dialogic modes of “speaking with” or “speaking alongside” [1]. This shift challenged hegemonic discourse and disrupted the one-dimensional representation of social groups. Ethnic minority documentaries, however, remain largely produced by outsiders, which often results in their subjects being represented as objects of anthropological observation. and presented as cultural specimens. In fact, voice-over narration, far from being a neutral tool, carries strong discursive power: it frames cultural sovereignty, reshapes identities, and redistributes the right to narrate. The purpose of this paper is to investigate how voice-over narration redistributes narrative rights through linguistic strategies in documentaries focusing on ethnic minority themes. Currently, there are numerous related studies that mainly criticize the uneven distribution of the right to speak

in ethnographic images from cultural-political and visual image perspectives, and explore the status of the narrator in different modes of narration in documentaries. Some linguistic research also connects media discourse to ideology and social power. However, there remains a scarcity of empirical research on the operation of voice-over narration itself—how its language constructs minority images and affects the distribution of narrative rights. This study addresses this gap by applying discourse analysis to the Chinese documentary *Image Journal of Fifteen Endemic Minorities in Yunnan, China*. It investigates the linguistic strategies of narration and examines how they shape the subject's right to narrate. The goal is to propose a more ethical framework for voice-over narration in documentaries about ethnic minorities.

2. Linguistic strategies: constructing the image of ethnic minorities

2.1. Psychological distance brought by third-person perspective narration

In the field of film studies, voice-over is defined as the voice of an off-screen narrator, and in documentaries, it is usually the commentator who provides a third-person narration that orients the audience to understand the content of the image [2]. In most situations, the creators did not aim to emphasize the power differentials between ethnic groups and to marginalise ethnic minorities. However, the third-person pronouns, such as "he", "they", and "this ethnic group", appear in the narration of the voice-over, which still isolates the audience from the subjects on the screen at a psychological level.

This isolation effect can be evidenced by neurolinguistic studies. Hartung et al. [3] demonstrated this through behavioural experiments and physiological data measurements. The results show that the use of third-person pronouns in reading texts significantly reduces readers' engagement, immersion, and appreciation of the story, essentially widening the psychological distance between the narration and the audience. The results of this experiment reveal the general effect of third-person narration on psychological distance, an effect that can also be extended to the field of film from text, where third-person narration in documentaries similarly diminishes the audience's sense of immersion, increases the understanding gap between the audience and the expression, and makes it difficult for them to truly enter the world of minority life. In this case, viewers are often unable to enter the narrated world and experience it alongside the characters, but rather sit calmly off-screen and observe the lives and cultures of "this ethnic group."

Relevant findings in translation studies can further corroborate this mechanism. In film and television translation, a first-person narration that retains the original voice can maximize the translator's invisibility. Conversely, if a third-person voice-over is used, it inserts an additional 'interpreter' between the audience and the character, thus altering the audience's reception [4]. Applied to the documentary context, this means that when the behavioural actions of ethnic minority members are transformed into third-person narration, the audience perceives not 'their own narrative' but an irrelevant person that 'speaks for them'.

The inequality of power in the image-making process makes it a challenging task to accurately portray the "other" [5]. During the creation process, producers often unconsciously embed their own stance in the narrative, and third-person narration is a concrete embodiment of this stance. In minority documentaries, this phenomenon is not only a "linguistic distance" but also an externalization of "power distance." This kind of narration method provides seemingly accurate information and rich cultural background on the surface, but in fact, it erects a "transparent glass wall" between the audience and the subjects, so that the audience can clearly observe their stories,

but cannot participate in their inner world activities, preventing real dialogue and resonance. In a sense, this reveals the dilemma of distributing speaking rights in minority documentary films.

2.2. The compulsory power of grammatical modifiers

In ethnic minority documentaries, one of the easiest ways for the producers to enhance the persuasiveness and authority of the film is to add a large number of modal adverbs in the voice-over narration, such as "must", "should", etc. These words are known as deontic modality in linguistics, and their core function is to emphasise the necessity of an action, generally depending on some kind of authority, often the speaker [6]. These words are not just grammatical modifiers, but a tool that directly and forcefully guides the audience's objective judgments and subjective feelings. For example, in the documentary film *Images of Fifteen Endemic Ethnic Minorities in Yunnan, China*, the narrator describes an activity of the Hani festival as follows: "As one of the leaders, Ma Xueguang must buy a cow that meets the requirements before the festival arrives." When the narrator uses this kind of language repeatedly, it undoubtedly establishes an atmosphere of unquestionable 'truth' for the film, which means that there are no other possibilities or options for the characters in the film.

In addition to the modal adverbs, there is another adverb we should be wary of: always. For example, in this documentary, the narrator describes the people of the Jino ethnic group by saying, "In the blood of the Jinuo people, music has always been closely associated with life and has never left." Such absolute expressions as "always" and "never" can easily leave the audience with the impression that "all Jinuos love music" without exception. Therefore, the audience may overlook the existence of individuals who are not skilled at music or do not enjoy it, and thus interpret a diverse group as a single "musical people." The vocabulary can manipulate the presentation of information and implicitly guide the audience's attitudes and perceptions of the film's content [7]. This is the power of this frequency adverbs: by simplifying and summarising, it compresses the complexity of reality into a unified image that is easy to accept.

The voice-over in early documentary films often appeared as an authoritative voice-of-God, frequently directed at the audience with an admonishing and instructive tone [8]. This voice creates an "educated" state of mind in the audience, as if they are listening to an omniscient narrator telling a story about the Other. In fact, the omniscient narrator is also unreliable, as producers may intentionally mislead the narrative for some purpose [9]. In the above example, we can speculate that the omniscient narration plays a role similar to that of a "government propaganda film", whose purpose is to guide the audience into a certain path of understanding. To meet this purpose, the producer removes the individual voice of ethnic minorities, which on the surface enhances the authority of the film and the consistency of the narrative, but it also means that the viewer is forced to accept a single interpretation and lacks the space for questioning and dialogue.

Therefore, the frequent use of those adverbs in ethnic minority documentaries is not an accident on the linguistic level, but is deeply involved in the construction of power relations. On the one hand, it helps producers establish a "scientific, objective and authoritative" image, but on the other hand, it simplifies and absolutizes ethnic minorities, obscuring the real and diverse individual voices.

3. Evaluative language and cultural reproduction: how narrators manipulate audience understanding

3.1. Evaluative adjectives and image construction

To make ethnic minority documentaries more romantic and mysterious, the rhetorical strategies in the voice-over often employ adjectives with emotional connotations to construct an image of the group, such as "mysterious", "ancient", and "pure", as well as other abstract adjectives that convey a seemingly beautifying effect. On the surface, these adjectives seem to imply admiration for a certain group of people and nostalgia for a past life that has not been corrupted by modern civilization, but in essence, they are expressions of superiority, which can easily lead to the formation of "the noble savage stereotype" [10]. It allows minority groups to be portrayed as alien to modern civilisation, remaining primitive and unstained.

This stereotype firstly comes from the deliberate use of 'primitivisation' [11], whereby ethnic groups are often portrayed as backward and homogenous in their lifestyles. Due to the influence of mainstream film and media, indigenous groups are often labeled as "stuck in the past" [12], which completely ignores their contemporary subjectivity and future creativity. At the same time, this phenomenon also implies the inherent perspective of mainstream ethnic groups' production teams, which define themselves as "modern", "advanced", "sophisticated", a hidden logic of antithesis that reinforces "the savage stereotype" and implies a comparison of high and low cultural orders.

Another strategy to counter "the noble stereotype" in Brylla's text is the "romanticisation". Under this strategy, ethnic groups are portrayed as the kind of people who dance with the gods, live with nature, and have never been persecuted by the outside world. In most cases, the viewer has no way to know whether the "primitive people" use modern objects and live modern lives, because this information will allow the audience to have something in common with the "primitive people", shorten the psychological distance, and thus make their pure and almost sacred image lose its viewability on the screen. Therefore, the presence of modern elements in a film is often seen as a threat to the idyllic traditional life [10].

Based on the theory of linguistic intergroup bias, linguists have found that people tend to describe stereotypes in abstract terms and counter-stereotypes in concrete terms [13]. This explains why narrators in documentaries frequently use abstract words such as "ancient", "mysterious", and "pure", which not only convey emotional colors quickly but also reinforce stereotypes in a subtle way. They not only convey emotion quickly but also subtly consolidate stereotypes. Taking the documentary "Images of Fifteen Unique Ethnic Minorities in Yunnan, China" as an example, the narrator, when describing the blessing ceremony of the Blang ethnic group, uses the adjective "holy" to depict the cotton threads tied around the wrists; and portrays the Blang ethnic villages as "hidden in the depths of Bada Mountain" and "continuing the ancient and pure folk customs"; the Lahu musical instrument, Lusheng, whose tone is characterized as "ancient and pure, and full of remembrance and veneration of ancestors"; while the description of the Hani girl is "primitive in style". These adjectives seem to give a unique charm to ethnic cultures, but in fact, they are a kind of "lazy labelling" which compresses ethnic cultures into a highly generalized symbol. This fixed template label of "noble savage" is directly attached to countless complex individuals. This kind of rhetorical language lurks in the danger of symbolising the image of ethnic minorities, influencing the audience's emotional attitudes towards the characters and manipulating their understanding at the same time. What the audience ultimately sees is not a genuine, comprehensive image, but one constructed through rhetoric that employs stereotypes, revealing the creator's arrogant attitude.

3.2. Absence of original sound and silence of subjects

In most countries, the production teams of ethnic minority videos are mostly from the mainstream ethnic groups, and they have already seized most of the discourse power in the process of planning, filming and editing, so the contents presented in such a discursive framework are more about "their stories" endorsed by the production teams than "our stories" told by the subjects themselves. Therefore, if the narration in the film relies mainly on the voice-over rather than the voices of the ethnic minorities themselves, their space to speak will be further compressed or even overridden. Such an override not only means that the audience cannot hear the subject's mother tongue expression, but also deeply reflects the unequal distribution of discourse power in the film.

To explore the relationship between the rate of voice-over narration and discourse power, this study quantitatively analyzes the composition of voices in documentaries. Taking the first episode of the first season of the 2022 Chinese documentary "Images of Fifteen Endemic Ethnic Minorities in Yunnan, China" as an example, the researcher cut off the opening and closing credits, exported the subtitle file to SRT format and then converted it to CSV format, imported it into Excel, and then classified and annotated the subtitles one by one. Finally, sum up the total duration of narration, dialogue, and other sounds respectively.

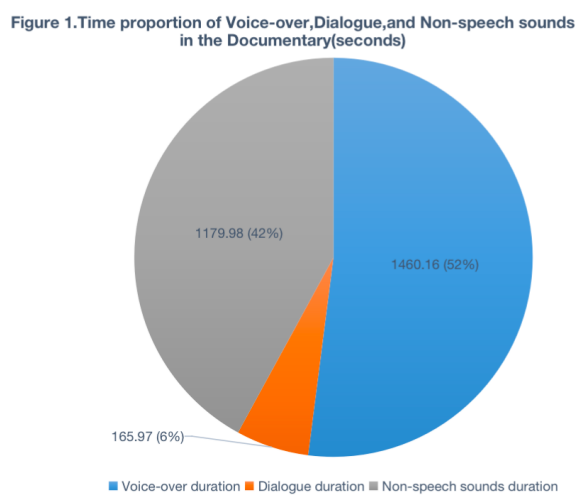


Figure 1. Time proportion of voice-over, dialogue, and non-speech sounds in documentary (seconds)

The data show that the Chinese (the dominant national language in China) voice-over narration accounts for more than half of the total film time, reaching 24 minutes out of 47. The total duration of the dialogue (including audio-visual synchronized dialogue and monologue) is less than three minutes, of which the Hani language accounts for only one-third of the dialogue, approximately 1 minute. That is to say, in this 47-minute documentary on ethnic minorities, the adoption of these ethnic languages in the narration accounts for only 1 minute, or less than 2%. Even if there are occasional dialogue segments, they are often short sound bites added to complement the narrator's narration, and they lack the presentation of the complex identities and sophisticated emotions of the individuals. This result shows that the native language is seriously absent in this documentary. This phenomenon is not only a deliberate documentary strategy but also an objective representation of the unequal distribution of discourse power.

In discourse analysis studies, "silence" and "absence" do not simply refer to the absence of voice or information, but are closely related to the relationship of power. "Silence" has been defined as

"the suppression or non-appearance of expected voices" [14]. When viewers watch documentaries, they naturally expect to hear the realistic voices of the ethnic minorities themselves, to understand their true inner activities and emotions, but the reality is that the production team compresses or even deletes their speeches during the editing process, and replaces them with the narrator's exquisite and literary third-person voice-over narration. For example, in the episode "Lahu", the narrator takes the place of the individual: "For them, it seems that only this place is their real home"; in the description of a couple from the Brown tribe, the narrator speaks directly without context: "Their love of music and their simple original intentions are the greatest motivation for them to keep playing and singing". The frequent occurrence of such a narrative suggests that this is a deliberate creative strategy chosen by the director. This strategy presents the subjective feelings of the subjects directly in the voice of the professional narrator, masking the original sound. This "speak for" strategy is considered the most hegemonic of the four narrative relationships in documentary film [1], resulting in the suppression of the subject's self-representation and autonomous expression. Further, this "speak for" strategy may come from two situations: either as a paraphrase based on interviews, or as a subjective speculation and choreography by the directing team. In either case, the producers did not provide space for ethnic minorities to actively express their real thoughts, nor did they encourage them to speak out in their mother tongue, thus undermining the opportunity for ethnic minority groups to reclaim their identities. This "silence" does not refer to a physical mute, but rather to the loss of the subject's right to express themselves and their helplessness to be "spoken for" by an authoritative voice. "Absence" is more complete than "silence", meaning that certain voices are never even considered in the text or in the construction of the narrative [14]. For example, in the planning stage of the documentary, the director did not consider national language and individual voices as necessary elements, and then naturally eliminated the traces of their voices, so that the audience would unconsciously ignore the missing parts during the watching process, and habitually accept the single perspective of the narrative. Indeed, who gets to have a voice and who is ignored is essentially the result of the operation of power [15]. This perspective can directly help us understand why the proportion of original voices in minority documentaries is low: it is not a matter of limited filming resources or technology, but rather the narrative preferences and logic of rights that have been developed by cultural producers over time. The "silence" and "absence" of original sound produced by this preference and logic can be regarded as a kind of cultural violence: it deprives ethnic minority individuals of their right to narrate and the audience of their right to access authentic cultural information.

4. Conclusion

This study takes the Chinese documentary *Fifteen Endemic Ethnic Minority Video Documentaries* in Yunnan, China, as an example and conducts a discourse analysis on its narrative text. The findings reveal that the voice-over narration strategy is a significant manifestation of cultural power relations in minority documentaries. The third-person narration, modifying adverbs and abstract adjectives, and other linguistic forms used in the strategy not only create a stereotypical image of the ethnic group but also strengthen the psychological distance between the viewer and the subject, ultimately leading to the ethnic group being viewed as a "sample". The phenomenon of the narrator overriding the original soundtrack also reflects the centralisation and imbalance of discourse power in the production process.

This study advocates for the application of discourse analysis in examining documentary voice-over narration, highlighting the importance of balancing the narrative power of language, which is crucial to understanding the creation of ethnic minority documentaries and their cultural

dissemination. From the theoretical level, this study combines discourse analysis research and documentary narrative critique, enriching the linguistic perspective of documentary research and providing new empirical evidence for exploring the power relationship between image and voice. On the practical level, this study emphasises the need for greater ethical sensitivity and narrative pluralism in the creation of ethnic minority videos, so that ethnic minority groups can move from being passive "objects" to active "subjects" of expression. A more inclusive narrative strategy not only helps promote the understanding of viewers from different cultural backgrounds but also corrects the invisible violence in linguistic narratives, promoting the creation of films and TV programs that respect multicultural expressions while achieving a higher level of authenticity and fairness.

References

- [1] Ruby, J. (1991). Speaking for, speaking about, speaking with, or speaking alongside an anthropological and documentary dilemma. *Visual Anthropology Review*, 7(2), 50-67. <https://doi.org/10.1525/var.1991.7.2.50>
- [2] Kuhn, A., & Westwell, G. (2012). *A dictionary of film studies*. Oxford University Press. <https://doi.org/10.1093/acref/9780199587261.001.0001>
- [3] Hartung, F., Burke, M., Hagoort, P., & Willems, R. M. (2016). Taking perspective: Personal pronouns affect experiential aspects of literary reading. *PloS One*, 11(5), e0154732. <https://doi.org/10.1371/journal.pone.0154732>
- [4] Matamala, A. (2018). Voice-over: Practice, research and future prospects. In *The Routledge handbook of audiovisual translation* (pp. 64-81). Routledge.
- [5] Tang, R., & Zou, Y. (2024). Language, cultural identity, and spirituality—A case study of the representation of minority ethnic subjects in chinese ethnographic documentary. *Studies in Ethnicity and Nationalism*, 25(2), 180-199. <https://doi.org/10.1111/sena.12447>
- [6] Palmer, F. R. (2001). *Mood and modality*. Cambridge University Press. <https://doi.org/10.1017/CBO9781139167178>
- [7] Fairclough, N. (2013). *Language and power*. Routledge.
- [8] Nichols, B. (2024). *Introduction to documentary*. Indiana University Press. <https://doi.org/10.2307/j.ctt2005t6j>
- [9] Wiener, J. (2007). The omniscient narrator and the unreliable narrator: The case of atomic café. *Film & History: An Interdisciplinary Journal of Film and Television Studies*, 37(1), 73-76. <https://doi.org/10.1353/flm.2007.0037>
- [10] Brylla, C. (2023). *Documentary and stereotypes: Reducing stigma through factual media*. Springer Nature Switzerland.
- [11] Shankar, A. (2020). *Primitivism and race in ethnographic film: A decolonial re-visioning*. Oxford University Press. <https://doi.org/10.1093/obo/9780199766567-0245>
- [12] Gergaud, S. (2021). Indigenous cinemas and futurisms: How Indigenous visual arts are shifting the narratives and offering a new relationship to space and landscapes. *Recherches Anglaises Et Nord-américaines*, (54), 151-169. <https://doi.org/10.4000/ranam.862>
- [13] Maass, A., Salvi, D., Arcuri, L., & Semin, G. R. (1989). Language use in intergroup contexts: The linguistic intergroup bias. *Journal of Personality and social Psychology*, 57(6), 981. <https://doi.org/10.1037//0022-3514.57.6.981>
- [14] Schröter, M., & Taylor, C.. (2017). *Exploring silence and absence in discourse: Empirical approaches*. Springer.
- [15] Achino-Loeb, M. L. (2005). *Silence: The currency of power*. Berghahn Books. <https://doi.org/10.3167/9781845451301>