

A Review of the Current Situation of Henan External Publicity Video Translation and Research on the Construction of a Multimodal Framework

Junye Li, Jia Liu

School of Culture & Education, Shaanxi University of Science and Technology, Xi'an, China

Email: 445368986@qq.com

How to cite this paper: Li, J. Y., & Liu, J. (2026). A Review of the Current Situation of Henan External Publicity Video Translation and Research on the Construction of a Multimodal Framework. *Open Journal of Social Sciences*, 14, 640-658.
<https://doi.org/10.4236/jss.2026.148038>

Received: July 10, 2026

Accepted: August 23, 2026

Published: August 26, 2026

Copyright © 2026 by author(s) and Scientific Research Publishing 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

Abstract

This study focuses on multimodal translation practices in Henan's international publicity videos to identify prominent cross-cultural communication challenges and propose corresponding optimization strategies. Driven by advances in digital media, video has emerged as a vital medium for disseminating regional culture to global audiences. Against this digital backdrop, international publicity videos have evolved beyond monolingual transmission toward integrated multimodal communication; their translation thus entails far more than linguistic transfer, encompassing cultural recoding and the reconstruction of visual meaning. Nevertheless, existing translation practices for Henan's promotional videos are plagued by prevalent defects including cultural misreading, oversimplified semantics, and poor synergy between audio and visual components, which undermine overall communication outcomes. To address these drawbacks, this paper develops the STM (Skopos-Target-Multi) multimodal collaborative translation framework by synthesizing Skopos Theory, Hofstede's Cultural Dimensions Theory, and Multimodal Discourse Analysis (MDA). Centered on communicative skopos, the framework adopts cultural discount assessment as an intermediate analytical step, realizes meaning delivery via multimodal synergy, and enables iterative dynamic adjustment through a built-in feedback loop. Case analyses of official Henan promotional videos illustrate how the STM framework may mitigate cultural misinterpretation, align multimodal information, and facilitate smoother cross-cultural communication.

Keywords

Skopos Theory, Cultural Discount, Multimodal Discourse Analysis,

1. Introduction

Rapid advances in global digital media technologies have made video a pivotal medium for cross-cultural communication and the construction of national and regional identities (Ren, 2023). Against this backdrop, Henan Province has launched cultural initiatives including the Translate Henan project, leveraging multilingual international publicity videos to present its rich historical heritage and local cultural identity. Such videos serve as a vital vehicle for promoting national and regional images worldwide.

Nevertheless, real-world translation practices for Henan's publicity videos face prominent obstacles: awkward rendering of culture-loaded terms, poor adaptability to cross-cultural contexts, and inadequate synergy between diverse audio-visual modal components (Hu & Li, 2024). These issues hinder audience comprehension and undermine the overall impact of cultural transmission.

Most prior scholarship adopts a single-theory lens. Typical examples include research on translation strategies guided by Skopos Theory, or applications of Multimodal Discourse Analysis to video communication (Chen et al., 2020; Wu, 2021). Few studies, however, systematically combine translational skopos, cross-cultural cognitive disparities, and multimodal coordination mechanisms within a unified analytical framework.

To fill this research gap, this paper puts forward the STM (Skopos-Target-Multi) multimodal collaborative translation framework. By synthesizing Skopos Theory, Cultural Dimensions Theory and Multimodal Discourse Analysis, this systematic analytical model is tailored to the translation of Henan's international publicity videos, with the ultimate goal of boosting cross-cultural communication performance.

This paper addresses three core research questions as follows:

- 1) What typical cross-cultural communication barriers emerge in the translation of Henan's international publicity videos?
- 2) How can the STM multimodal collaborative translation framework be constructed to interpret and resolve the aforementioned challenges?
- 3) How may the STM framework refine cross-cultural communication performance when applied to real-world case analyses?

2. Literature Review

2.1. Studies on Skopos Theory

First proposed by Vermeer (2004), Skopos Theory argues that translation activities should prioritize intended communicative skopos rather than rigid linguistic formal equivalence. Building on this foundation, Nord (1997) put forward the "function plus loyalty" principle, laying a more systematic theoretical foundation for publicity-oriented translation. Over the past decade, this theory has been

widely adopted in audiovisual translation and publicity text research, with scholars acknowledging its strong guiding value for skopos-driven translation practice (Hu & Li, 2024).

Nevertheless, practical applications of Skopos Theory tend to focus solely on linguistic translation strategies while neglecting the synergistic relations among visuals, soundtracks and subtitles within multimodal texts (Pérez-González, 2014). As a result, the theory fails to fully account for intricate cross-cultural dynamics in video communication and exhibits limited explanatory power for multimodal contexts. In particular, few systematic analyses have examined the coordination mechanisms between visual and verbal signifiers in video translation.

2.2. Cultural Dimensions Theory and Translation Research

Hofstede's Cultural Dimensions Theory serves as a vital analytical tool for cross-cultural communication. It unpacks the origins of cultural disparities through six core dimensions, including power distance, individualism-collectivism, and uncertainty avoidance.

Within translation studies, this framework is frequently deployed to interpret the handling of culture-loaded terms and cross-cultural conflicts. Even so, extant scholarship merely identifies cultural differences without establishing systematic workflows to translate cultural diagnostic findings into actionable multimodal translation strategies. Recently, Yuan and Zhu (2025) investigated cultural discount effects in audiovisual products from the dimensions of audience perception and communicative context, offering a new analytical lens for integrating Cultural Dimensions Theory with translation practice.

2.3. Multimodal Discourse Analysis Research

Multimodal Discourse Analysis (MDA) posits that meaning construction relies not merely on verbal language, but on the joint operation of multiple semiotic resources such as images, audio and subtitles (Zhang, 2009; Hu, 2007; Kress & van Leeuwen, 2001).

In audiovisual and publicity video translation research, MDA has been applied to unpack the meaning-making mechanisms of audio-visual semiotic resources (Chen et al., 2020; Wu, 2021). Deng (2020) systematically explored modal conflicts and compensatory strategies in audiovisual translation, while Wu & Liu (2021) adopted Beautiful China as a case to examine how multimodal translation shapes national image construction. However, existing MDA research remains largely descriptive, failing to link multimodal analytical outcomes to concrete translation decision-making procedures and thus weakening its practical instructional value.

2.4. Domestic and International Research at the Intersection of Cultural Tourism Publicity and Multimodality

The preceding sections have reviewed three core theoretical strands in translation studies: Skopos Theory, Cultural Dimensions Theory and Multimodal Discourse Analysis. Yet the applicability of these integrated theories requires empirical test-

ing within localized regional publicity contexts. In recent years, global and Chinese scholars have conducted abundant empirical studies on cultural tourism publicity translation and multimodal communication, providing practical references for the framework developed in this paper. This section reviews domestic provincial publicity translation research and cutting-edge international multimodal studies, before identifying prevailing research gaps and clarifying the paper's research entry point.

2.4.1. Domestic Research on Provincial Cultural Tourism Publicity Translation

This study selects Shandong, Shaanxi and Sichuan Provinces as comparative research cases. All three regions boast abundant historical cultural heritage, whose cultural characteristics and resource endowments are comparable to Henan's Central Plains culture. Existing studies of these provinces share identical methodological limitations, which reveal universal flaws in domestic regional publicity translation research. Specifically, Shandong-focused studies draw on scenic signs and printed brochures as primary corpus, relying exclusively on either Skopos Theory or Eco-translatology. Shaanxi's cultural relic translation research centers on static image-text annotations, while Sichuan's scholarship mainly analyzes stationary urban tourism slogans; the handful of short video studies only briefly describe superficial image-text correspondences. Collectively, research across these three provinces relies on static text corpus, lacks multimodal coordination analysis for short videos, and fails to trace cultural discount roots via Cultural Dimensions Theory—these constitute the common limitations of domestic regional publicity translation research.

2.4.2. International Cutting-Edge Research on Multimodality and Cultural Discount

The concept of cultural discount was first coined by Hoskins & Mirus (1988) in research on cross-border television program trade. They argued that cultural content originating from one culture loses appeal and communicative effectiveness when introduced to foreign target markets due to cultural gaps. This concept has since been widely adopted in translation studies to explain semantic loss when transferring culture-loaded elements across linguistic systems. Yuan and Zhu (2025) further confirmed that audiovisual works suffer prominent cultural discount effects during cross-cultural transmission, stemming from gaps in audience perception and communicative contexts.

Audiovisual translation has witnessed rapid theoretical advancement in international academia, with scholars exploring multimodal translation strategies, methodological shifts, and comparisons between human-generated and AI subtitles (Díaz Cintas & Remael, 2021). Nevertheless, no established specialized analytical framework has been developed for promotional videos featuring Chinese local cultural narratives.

In summary, while prior research has analyzed translation phenomena through

Skopos Theory, Cultural Dimensions Theory and Multimodal Discourse Analysis respectively, three prominent research gaps remain. First, single-theory frameworks cannot fully explain the intertwined “language-culture-multimodality” communication dynamics embedded in publicity videos. Second, few systematic transformation pathways connect cultural analysis results to tangible multimodal translation strategies. Third, multimodal analytical findings are rarely incorporated into formal translation decision-making workflows.

To fill these research gaps, this paper constructs the STM (Skopos-Target-Multi) multimodal collaborative translation framework. By synthesizing the three aforementioned theories, this framework establishes a complete logical chain spanning translational skopos formulation, cultural risk identification, and multimodal meaning delivery.

3. Research Methods

The literature review above demonstrates that Skopos Theory, Cultural Dimensions Theory and Multimodal Discourse Analysis have each yielded abundant research outcomes, yet few studies organically combine the three theories to address practical demands for regional international publicity video translation. Against this backdrop, this chapter elaborates the development and application of the STM multimodal collaborative translation framework from three sequential dimensions: corpus selection, research methodologies, and the logical design of the analytical framework.

The corpus of this study comprises 12 short and full-length official promotional videos released by the Henan Provincial Department of Culture and Tourism and Henan Provincial Government official platforms between 2020 and 2025. All materials contain complete visuals, voiceovers and bilingual subtitles, which bear typical multimodal features and cross-cultural communication attributes.

3.1. Corpus Selection Details

Four screening criteria define the 12-video corpus:

- 1) Authority: Videos are sourced exclusively from The People’s Government of Henan Province official website (<https://www.henan.gov.cn/>), excluding non-official clips.
- 2) Time range: Released 2020-2025, covering the Translate Henan campaign.
- 3) Multimodal integrity: Complete footage, Chinese voiceover and standard Chinese-English subtitles.
- 4) Cultural diversity: Cover Central Plains heritage, kung fu, ancestor culture and local industries.

Full 12-video inventory (original source: <https://www.henan.gov.cn/>; some re-distributed overseas)

- 1) Meet China, Hometown Henan, bilingual, 2021
- 2) Discover Henan, bilingual, 2022
- 3) Center of Heaven and Earth: Dengfeng World Cultural Heritage, bilingual,

2023

- 4) Shaolin: Zen and Kungfu, bilingual, 2020
- 5) Yellow Emperor Reimagined by AI, bilingual, 2024
- 6) The Legend of Youxiong: Ancestor Civilization of Henan, bilingual, 2024
- 7) Song Dynasty Stone Carvings along the Yellow River Levee, bilingual, 2023
- 8) Made in Henan: High-quality Industrial Brands, bilingual, 2025
- 9) The Ancient Capital Luoyang: A Millennial Civilization, bilingual, 2021
- 10) Kaifeng: Splendor of the Northern Song Dynasty, bilingual, 2022
- 11) The Grand Canal in Henan: Cultural Heritage Corridor, bilingual, 2025
- 12) Traditional Handicrafts of the Central Plains, bilingual, 2020

Rationale for six analysed extracts

Six clips are selected for three reasons:

- 1) Representativeness: Each stands for one distinct type of Central Plains cultural text in the corpus.
- 2) Typical translation defects: All contain prominent cross-cultural misreading and audio-visual modal mismatch.
- 3) Analytical value: Their flawed multimodal subtitles enable close examination of how the STM framework may refine translation outputs.

3.2. Operational Definition & Assessment Indicators of Core Concepts

Before case analysis, this study sets unified measurable standards for three core concepts:

- 1) Cultural discount Identification standard: Cultural elements in subtitles trigger foreign audience cognitive bias, semantic distortion or weakened cultural recognition, judged via Hofstede's cultural dimension gaps. Assessment index: High discount (literal translation misrepresents cultural symbols); Medium discount (culture-specific meaning simplified); Low discount (cross-cultural comprehension unobstructed).
- 2) Multimodal synergy Identification standard: Coordination degree among subtitles, voiceover and frame images in conveying identical cultural meaning. Assessment index: Optimal synergy (three modes complement each other); General synergy (partial modal separation); Poor synergy (text-image semantic conflict).
- 3) Communication effectiveness Identification standard: Whether the video accurately delivers Henan's regional culture and achieves overseas audience recognition. Assessment index: Cognitive layer (cultural information intelligibility); Affective layer (cultural identity); Behavioral layer (willingness to understand local culture further).

Reproducible Case Selection & Cultural Risk Coding Procedure.

3.3. Representative Case Screening Procedure

Step 1: Filter all 12 videos against the four corpus inclusion criteria laid out above.
 Step 2: Tag each video by its core Central Plains cultural theme (heritage, martial arts, ancestor culture, ancient relics, modern industry, intangible heritage).

Step 3: Extract one 30 - 90 second segment per distinct theme that contains prominent culture-laden subtitles and obvious multimodal mismatch, forming six representative analysis extracts.

3.4. Cultural Risk Coding Procedure

Three-tier deductive coding scheme is applied:

Tier 1: Identify culture-specific vocabulary that may cause cross-cultural misunderstanding;

Tier 2: Classify risks into three grades of cultural discount (low/medium/high) per the operational definition in Section 2;

Tier 3: Record mismatches among subtitle text, voiceover and visual frames to assess multimodal synergy. All coded risks are logged in a unified excel coding form for traceability.

3.5. Multi-Researcher Cross-Review Protocol

Two researchers independently screened candidate clips and completed cultural risk coding separately. Discrepancies in case selection or coding results were discussed to reach a consensus before formal analysis, ensuring the reliability of qualitative judgement.

This research adopts two analytical approaches: case study and comparative analysis. Case study is utilized to dissect representative translation samples from Henan's publicity videos under the three STM dimensions of Skopos, Target and Multi. Comparative analysis contrasts cultural conveyance and multimodal presentation between original subtitles and revised translations. Horizontal comparisons across diverse cases further reveal regular patterns of rendering culture-laden expressions within multimodal contexts.

The STM (Skopos-Target-Multi) multimodal collaborative translation framework encompasses three core tiers:

- 1) Skopos Tier: Clarify communicative objectives and audience positioning to set the overarching orientation for translation strategies;
- 2) Target Tier: Detect culture-laden elements and potential cultural discounts so as to pinpoint cross-cultural communication barriers;
- 3) Multi Tier: Integrate linguistic, visual and auditory semiotic resources to realize multimodal synergy and semantic compensation.

Furthermore, the model incorporates a feedback loop. Translation strategies are continuously refined through assessments of communication effectiveness to achieve dynamic optimization. The framework serves dual functions: analytical interpretation and actionable strategic guidance.

4. Construction of the STM Multimodal Collaborative Translation Framework

4.1. Theoretical Foundations of the STM Framework

When studying publicity video translation, any single theoretical lens fails to re-

solve complex cross-cultural communication dilemmas. Skopos Theory prioritizes the skopos of translation practice but lacks instruments for identifying cultural discrepancies. Cultural Dimensions Theory pinpoints the roots of cultural conflicts yet offers no direct actionable translation guidelines. Multimodal Discourse Analysis unpacks the generation of audiovisual meaning, but fails to establish systematic links with translation decision-making (Díaz Cintas & Remael, 2021).

Given these theoretical limitations, this paper integrates the three paradigms to build the STM multimodal collaborative translation framework, which forms a complete logical chain: translational skopos → cultural risk identification → multimodal delivery. The integration is not a simple overlay of theories, but a layered reconstruction aligned with the full cross-cultural communication workflow.

4.2. Overall Architecture of the STM Framework

Rather than mechanically combining three separate theories, the STM framework is constructed as a progressive analytical model grounded in cross-cultural communication logic. Its operation relies on three interconnected tiers:

1) Skopos Tier (Communication Goal Setting) This tier prioritizes the clarification of translation's communicative skopos and target audience profiles. When working on promotional videos, translators formulate overall communication strategies based on the target group's cultural background and cognitive habits, which deliver normative guidance for subsequent translation choices.

2) Target Tier (Cultural Discount Identification Mechanism) Drawing on Cultural Dimensions Theory, this tier systematically examines culture-laden information in source texts. It screens historical allusions, regional cultural symbols and value-laden expressions that may trigger misreading or cultural discount, thereby identifying latent cross-cultural obstacles.

3) Multi Tier (Multimodal Implementation Path) Rooted in Multimodal Discourse Analysis, this tier unifies verbal language, images, captions and audio within integrated translation tactics. Cross-modal synergy facilitates semantic compensation and optimized communication outputs, rendering translations more intelligible and influential for target audiences (Deng, 2020; Wu & Liu, 2021).

4.3. Progressive Logical Chain of the Three STM Tiers

The STM framework is distinguished by its sequential, closed-loop logic instead of parallel independent modules. Its operating flow proceeds as follows: First, the Skopos Tier sets overarching communication goals and imposes directional constraints on translation activities. Second, the Target Tier filters source cultural information and locates core cross-cultural transformation challenges based on pre-defined communicative skopos. Third, the Multi Tier leverages linguistic and multimodal resources to implement targeted translation tactics according to findings from the prior two tiers.

The three tiers form a sequential logic: goal formulation → risk diagnosis → stra-

tegic execution, constructing a systematic decision-making workflow for cross-cultural translation. This progression enables the translation process to move from abstract objectives to concrete multimodal implementation.

4.4. Feedback Loop of the STM Framework

Different from conventional linear translation models, the STM framework embeds a feedback loop to support dynamic iteration. By evaluating audience comprehension, cultural acceptance and real communication feedback, researchers can adjust the communicative skopos defined in the Skopos Tier, and further optimize cultural identification tactics (Target Tier) and multimodal presentation forms (Multi Tier) in reverse.

This feedback mechanism expands the STM framework's utility beyond textual analysis to practical translational iteration. The closed-loop iterative design greatly improves the model's adaptability and instructional value for international publicity video translation.

4.5. Explanation of the STM Framework Diagram

Different from conventional linear translation models, the STM framework embeds a feedback loop to support dynamic iteration. By evaluating audience comprehension, cultural acceptance and real communication feedback, researchers can adjust the communicative skopos defined in the Skopos Tier, and further optimize cultural identification tactics (Target Tier) and multimodal presentation forms (Multi Tier) in reverse.

As illustrated in **Figure 1**, the proposed STM framework consists of three progressive tiers connected by a feedback loop: Skopos Tier: Define core communication objectives Target Tier: Identify potential cultural discounts Multi Tier: Deliver meanings through multimodal coordination Feedback Loop: Optimize the entire translation strategy.

This architecture delivers a full-process workflow spanning translational goal-setting, cultural analysis and multimodal realization, providing a systematic analytical tool for international publicity video translation.

5. Case Analysis

To validate the efficacy of the STM (Skopos-Target-Multi) multimodal collaborative translation framework, this study extracts representative culture-laden phrases from official promotional videos issued by the Henan Provincial People's Government and Henan Provincial Department of Culture and Tourism, with flagship productions Meet China, Hometown Henan and Discover Henan as core analytical samples. All cases are examined comprehensively across three dimensions: subtitle wording, visual context, and cross-cultural communication outcomes.

Every case is unpacked along the three tiers of the STM framework as follows:

- 1) Skopos Tier: Examine communicative goals and audience positioning.
- 2) Target Tier: Detect cultural discounts and cross-cultural obstacles.

3) Multi Tier: Analyze multimodal representations and corresponding optimization approaches.

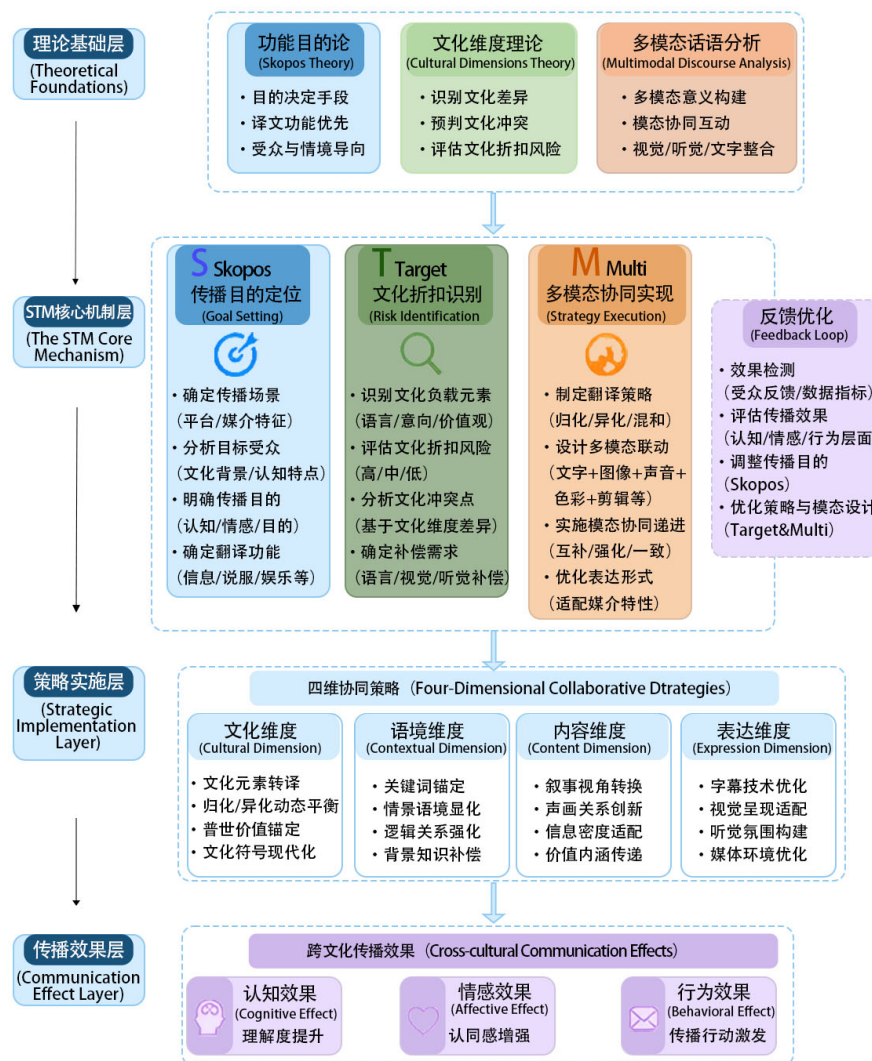


Figure 1. The STM multimodal collaborative translation framework.

5.1. Case 1



Figure 2. Visual-text collocation of the "Center of Heaven and Earth".

Timing: 00:25-00:35

Chinese source text: 这里是天地之中

Official English subtitle: Known as the “Center of Heaven and Earth”

STM revised subtitle: Known as the “Center of Heaven and Earth”, the cultural and cosmological heart of early Chinese civilization.

As illustrated in **Figure 2**, the phrase “Center of Heaven and Earth” originates from a promotional video featuring the historic architectural complex in Dengfeng, Henan, which bears the same name. The video adopts double-exposure footage blending a figure reaching toward the sunrise with an ancient observatory to visualize China’s ancient philosophical notion of the harmony between heaven and humanity.

From the Skopos Tier perspective, the video’s primary communicative goal is to interpret Dengfeng’s significance as an ancient hub integrating astronomical observation and ritual institutions, and to frame it as the geographical axis of Central Plains civilization embodied by the “Center of Heaven and Earth”. Accordingly, subtitles must balance recognizability for overseas audiences with accurate delivery of profound historical meanings.

From the Target Tier perspective, a verbatim rendering of this phrase tends to evoke Western biblical connotations, diluting its intrinsic geographical and ritual value and generating prominent cultural discount (Hoskins & Mirus, 1988). Drawing on Hofstede’s Cultural Dimensions Theory, Western societies shaped by individualism and low power distance lack the collective ritual cognition rooted in the harmony between heaven and humanity. This gap makes it hard for foreign viewers to grasp the spatial logic centered on imperial astronomical observation.

From the Multi Tier perspective, the footage already constructs a complete contextual backdrop of celestial observation, yet the literal subtitles fail to supplement implicit cultural meanings through visual cues. This separation between linguistic and visual semiotic resources undermines the narrative’s explanatory strength.

STM-Based Revised Translation: Known as the “Center of Heaven and Earth”, the cultural and cosmological heart of early Chinese civilization.

The supplementary attributive clause highlights the site’s dual identity as an astronomical and ritual landmark, aligning perfectly with the observatory visuals to realize complementary audio-visual meaning. This revised rendering preserves the unique cultural proper noun while embedding explanatory connotations, dispelling cross-cultural cognitive bias and delivering coordinated multimodal communication effects.

5.2. Case 2

Timing: 00:03-00:15

Chinese source text: 天下功夫出少林，少林寺是禅宗祖庭、功夫圣地，距今已有 1500 余年历史

Official English subtitle: With more than 1500 years of history, Shaolin Temple is the birthplace of Chinese Zen Buddhism and Chinese Kungfu.



Figure 3. Multimodal elements in the Shaolin Temple promotional clip.

STM revised subtitle: With a 1500-year legacy, Shaolin Temple stands as the sanctuary of Kungfu and the spiritual origin of Zen.

As illustrated in **Figure 3**, the original subtitle reads: With more than 1,500 years of history, Shaolin Temple is the birthplace of Chinese Zen Buddhism and Chinese Kungfu. This extract comes from the opening sequence of Henan's international promotional video. The footage features shots of the temple gate, ancient cypresses and bell tower to build an immersive Oriental Zen atmosphere, showcasing the distinctive Central Plains cultural icon where Zen practice and martial arts originate side by side.

From the Skopos Tier perspective, the video intends to position Shaolin Temple as a globally renowned cultural landmark. Its subtitles need to reflect the temple's 1500-year historical legacy and communicate the unique spiritual lineage shared by Zen and Kungfu, striking a balance between conciseness and profound spiritual connotation.

From the Target Tier perspective, the term “birthplace” only denotes geographical origin, eroding the ancestral lineage embodied by the temple and resulting in oversimplified semantics. Per Hofstede's Cultural Dimensions Theory, audiences from individualistic Western cultures struggle to identify with sect-based collective heritage. Additionally, groups with low uncertainty avoidance show weak acceptance of this integrated Zen-martial arts symbol, which intensifies cultural discount effects.

From the Multi Tier perspective, lengthy original subtitles occupy excessive screen time. The static wording of “birthplace” clashes with the fluid, tranquil visual imagery, breaking the harmony between audio and visual narration.

STM-Based Revised Translation: With a 1500-year legacy, Shaolin Temple stands as the sanctuary of Kungfu and the spiritual origin of Zen.

This revision highlights the temple's dual spiritual identity while shortening the sentence to fit subtitle time limits. It echoes the serene visual atmosphere shaped by ancient trees and bell towers, aligning linguistic and visual modalities to construct a cohesive narrative of Zen and martial arts and reduce comprehension barriers for international viewers.

5.3. Case 3



Figure 4. Title Design: AI 绘轩辕.

Timing: 00:03-00:08

Chinese source text: AI 绘轩辕

Official English subtitle: Yellow Emperor in AI's Depiction

STM revised subtitle: Yellow Emperor: Reimagined by AI

As illustrated in **Figure 4**, the expression “Yellow Emperor in AI’s Depiction” is treated in the translation as “Yellow Emperor in AI’s Depiction”, mainly derived from the opening title of the promotional video of Henan International Communication. The video begins with grand background music, golden light effects and tech-savvy dynamic visual design, aiming to present an innovative narrative of digital technology reconstructing the primogenitor culture of China.

On the Skopos level, the communication objective takes into account two points: one is to trace the roots of the Yellow Emperor civilization, and the other is to highlight the characteristics of AIGC digital creation. The title translation needs to balance the historical weight and the technological trend.

At the Target level, “AI’s Depiction” focuses on static depiction and fails to reflect the creative attributes of AI’s active generation and reconstruction of traditional culture. Combining Hofstede’s long-term/short-term orientation dimension, Western short-term orientation culture places more emphasis on immediate creative outcomes, neglects the long-term historical narrative of the ancestors, and weakens the depth of cultural dissemination.

At the Multi level, the opening is full of futuristic visual effects, but the static translation style is disconnected, and the language modality fails to echo the digital innovation image, weakening the visual impact at the beginning.

Based on STM framework optimized translation: Yellow Emperor: Reimagined by AI. precisely embody the creative logic of AI reshaping traditional culture, fit the technological visual expression of dynamic golden font, take into account the core of ancestral culture traceability and the theme of digital innovation, coordinate the three layers of modality of language, picture, brand, and strengthen the innovative positioning of the promotional video.

5.4. Case 4



Figure 5. Audio-visual combination of the Youxiong-themed scene.

Timing: 00:10-00:15

Chinese source text: 黄帝是有熊国少典之子

Official English subtitle: the Yellow Emperor was the son of Shaodan from the Youxiong State

STM revised subtitle: The Yellow Emperor descended from Shaodan of the ancient Youxiong clan

As illustrated in **Figure 5**, from the Skopos Tier perspective, this clip's core communicative goal is to interpret the Yellow Emperor's tribal ancestral roots, a foundational symbol of Central Plains origin culture. Subtitles must render obscure prehistoric clan terminology intelligible to international viewers with no background on early Chinese tribal societies.

From the Target Tier perspective, the literal rendering "Youxiong State" creates prominent medium-level cultural discount. Guided by Hofstede's cultural dimensions framework, Western audiences adopt fixed modern nation-state cognitive logic, and cannot distinguish primitive tribal confederations from formal sovereign states, resulting in misrecognition of Huaxia ancestral totem culture.

From the Multi Tier perspective, the visual sequence features primeval woodland, giant brown bear totems and primitive tribal characters, fully materializing the symbolic connotation of the "Bear Clan". The original subtitle wording conflicts with the prehistoric tribal visual cues; the revised translation substitutes "State" with "clan" to coordinate textual semantics and visual totem information, achieving consistent multimodal meaning delivery.

5.5. Case 5

Timing: 00:17-00:38

Chinese source text: 桥东侧保留宋代堤岸石壁，石壁上雕凿有海马、瑞兽、鹤禽、祥云浮雕纹饰

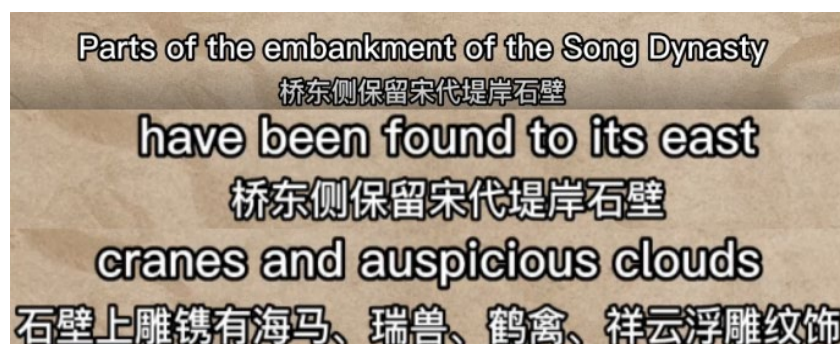


Figure 6. Subtitle-visual mismatch on Song-Dynasty stone reliefs.

Official English subtitle: Parts of the embankment of the Song Dynasty have been found to its east on which are reliefs of sea horses, auspicious beasts, cranes and auspicious clouds.

STM revised subtitle: Its east side retains Song Dynasty embankment stone walls carved with Tianma (mythical winged horses), auspicious beasts, cranes and sacred clouds.

As illustrated in **Figure 6**, the original subtitle reads: Parts of the embankment of the Song Dynasty have been found to its east on which are reliefs of sea horses, auspicious beasts, cranes and auspicious clouds. This excerpt is taken from a thematic short clip about Song Dynasty embankment stone carvings in Henan's international promotional video. Delicate golden outlines trace celestial horses, auspicious animals and cloud patterns, showcasing China's traditional totem aesthetic centered on blessings and good omens.

From the Skopos Tier perspective, this segment aims to introduce Song Dynasty visual arts and auspicious cultural symbolism to global audiences. The subtitles must clearly differentiate China's mythical sacred creatures from real-life Western marine species.

From the Target Tier perspective, the literal term "sea horses" incorrectly equates the mythical Tianma with tiny marine seahorses, leading to fundamental symbolic distortion. Per Hofstede's uncertainty avoidance dimension, Western audiences tend to categorize things in realistic, concrete terms and struggle to grasp abstract composite mythological totems. Such word-for-word translation erases the unique auspicious cultural implications embedded in Chinese iconography.

From the Multi Tier perspective, the footage displays grand stone reliefs of mythical creatures, yet the original text anchors these figures to ordinary marine life. This creates severe dissonance between audio and visual semiotic resources, breaking the coherent cultural narrative built around auspicious animal totems.

STM-Based Revised Translation: ... on which are reliefs of Tianma (mythical winged horses), auspicious beasts, cranes and sacred clouds.

This revision rectifies the symbolic misunderstanding by labeling the mythical horses as winged Tianma, while replacing plain "auspicious clouds" with "sacred clouds" to amplify auspicious undertones. The wording aligns with the Oriental aesthetic style of the stone carvings, resolves cognitive dissonance between visuals

and text, and fully delivers the profound cultural meanings of traditional Chinese totems.

5.6. Case 6

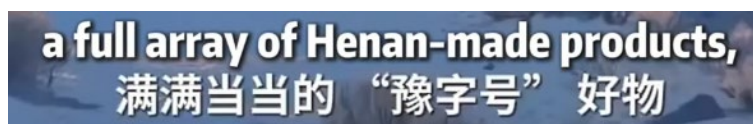


Figure 7. Multimodal presentation of Henan-made product footage.

Timing: 01:04-01:10

Chinese source text: “豫字号”好物

Official English subtitle: Henan-made products

STM revised subtitle: the signature quality brands of Henan.

As illustrated in **Figure 7**, the original subtitle segment reads: Henan-made products ... This clip comes from Henan’s international promotional video. The video adopts fast-paced split screens and rapid shot transitions, featuring montages of intelligent manufacturing and premium textiles to build a public brand identity for Henan’s high-quality regional industries.

From the Skopos Tier perspective, this segment is designed to promote Henan’s unified regional brand and demonstrate its modern manufacturing prowess. Subtitles therefore need to convey the core brand value embedded in regional industrial clusters and premium quality credentials.

From the Target Tier perspective, the plain phrase “Henan-made products” merely marks geographic origin and fails to reflect the value of Henan’s unified regional collective brand. Under Hofstede’s individualism-collectivism dimension, Western societies prioritize independent individual enterprises and struggle to comprehend region-wide collective industrial brands. This mismatch undermines brand dissemination and generates cultural discount effects.

From the Multi Tier perspective, the dense industrial footage demands subtitles that synthesize scattered visual cues. Generic origin-based phrasing cannot unify fragmented manufacturing scenes, leaving foreign audiences unable to form a cohesive impression of Henan’s premium manufacturing industry.

STM-Based Revised Translation: ... the signature quality brands of Henan.

The term “signature” emphasizes regional representativeness, while “quality brands” highlights the cluster’s premium positioning. This wording matches the fast-paced industrial visuals, enables subtitles to integrate dispersed visual information, forms a unified brand impression of Henan manufacturing among international viewers, and realizes coordinated multimodal communication between text and visuals.

6. Discussion

This paper constructs the STM (Skopos-Target-Multi) multimodal collaborative translation framework by synthesizing Skopos Theory, Cultural Dimensions The-

ory and Multimodal Discourse Analysis, establishing a complete systematic analytical workflow for international publicity video translation. Compared with conventional single-theory analytical lenses, the STM framework delivers three core strengths:

First, it reinforces the skopos-driven logic of translation practice by setting the Skopos Tier as the starting point of all translation decision-making, which breaks the confines of equivalence-centered traditional translation scholarship (Vermeer, 2004; Nord, 1997).

Second, the Target Tier embeds a cultural discount identification mechanism. It transforms static descriptive cultural dimension analysis into a dynamic process of diagnosing cross-cultural obstacles, and bridges the disconnect between cultural analysis findings and actionable translation strategies (Hoskins & Mirus, 1988; Yuan & Zhu, 2025).

Third, the Multi Tier integrates all semiotic resources including text, visuals and audio within one unified analytical system to realize cross-modal coordinated optimization (Zhang, 2009; Deng, 2020). This innovation expands translation research beyond pure textual analysis to cover the entire cross-cultural communication ecosystem. By synthesizing the three foundational theories, the STM framework shifts analytical logic from static textual description to dynamic mechanism interpretation.

Conventional Skopos Theory only addresses the rationale behind translation, yet fails to provide solutions for identifying cultural barriers and designing multimodal representations. Cultural Dimensions Theory can trace the root of cultural disparities but offers no concrete operational translation workflows. Meanwhile, Multimodal Discourse Analysis remains largely descriptive and cannot directly inform practical translation decision-making.

By contrast, the STM framework establishes a sequential “Skopos-Target-Multi” progressive mechanism that forms a closed logical loop: defining communicative skopos → identifying cultural barriers → implementing multimodal translation solutions (Chen et al., 2020; Wu & Liu, 2021). To a large extent, it remedies the practical application defects of the three individual theories mentioned above.

When applied to translation practice for Henan’s international publicity videos, the STM framework demonstrates potential to improve cross-cultural communication performance within the analysed cases. Its practical benefits fall into three categories: 1) Offer approaches to improve the intelligibility of culture-laden content and reduce risks of cultural misreading caused by word-for-word literal translation; 2) Optimize multimodal information synergy to align subtitles, visuals and verbal narration consistently; 3) Improve overall communication outcomes and adapt local cultural narratives to international audiences’ cognitive conventions.

Unlike traditional analytical models that prioritize the explanatory power of a single theory, the STM framework centers on the logical interplay between its three tiers. It thus carries both theoretical integrative value and robust practical instructional significance.

7. Conclusion

This study investigates multimodal translation challenges embedded in Henan's international publicity videos. The analysis reveals two pervasive flaws in their cross-cultural communication: prevalent cultural discount and semantic distortion, especially in the rendering of culture-laden terms; and poor coordination between the multiple semiotic modes of language, images and subtitles, which jointly undermine overall communication efficacy.

The proposed STM multimodal collaborative translation framework integrates Skopos Theory, Cultural Dimensions Theory and Multimodal Discourse Analysis. It organically unifies skopos orientation, cultural risk diagnosis and multimodal meaning delivery, delivering an operable systematic analytical model for international publicity video translation and enabling more standardized, structured translation practice.

Nevertheless, the framework has three notable limitations. First, the corpus scope is confined exclusively to Henan's promotional videos; follow-up research may expand the dataset to cultural communication materials from other provinces to validate its generalizability. Second, this study relies solely on qualitative case analysis without quantitative statistical support or systematic measurable indicators to quantify communication effects; all judgements of translational optimisation are derived from comparative textual and multimodal analysis rather than empirical audience data. Third, the framework's adaptability requires further testing and refinement across diverse cross-cultural communication contexts.

Subsequent research can advance this field by broadening corpus coverage, adopting quantitative analytical methods, and refining evaluation systems for multimodal communication performance.

Author Contributions

Conceptualization, J-Y. L.; methodology, J-Y. L. and J. L.; software, J. L.; validation, J-Y. L. and J. L.; formal analysis, J. L.; investigation, J. L.; resources, J-Y. L.; data curation, J. L.; writing—original draft preparation, J. L.; writing—review and editing, J-Y. L. and J. L.; visualization, J. L.; supervision, J-Y. L.; project administration, J-Y. L.; funding acquisition, J-Y. L. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- Chen, X., Pan, H., & Pan, L. (2020). Multimodal Shifts in Translation Studies: Current Status and Prospects. *Journal of Foreign Languages*, No. 2, 80-87. (In Chinese)
- Deng, Q. (2020). Modal Conflict and Compensation in Film and Television Translation. *China Translation*, 41, 45-52.
- Díaz Cintas, J., & Remael, A. (2021). *Audiovisual Translation* (2nd ed.). Routledge.

- Hoskins, C., & Mirus, R. (1988). Reasons for the U.S. Dominance of the International Trade in Television Programmes. *Media, Culture and Society*, 10, 525-542.
<https://doi.org/10.1177/016344388010004006>
- Hu, K., & Li, S. (2024). Characteristics and Translation Principles of Multimodal Diplomatic Discourse. *Journal of Beijing International Studies University*, 46, 14-26. (In Chinese)
- Hu, Z. (2007). Multimodalization in Social Semiotic Research. *Language Teaching and Research*, No. 1, 1-10. (In Chinese)
- Kress, G., & Van Leeuwen, T. (2001). *Multimodal Discourse*. Edward Arnold.
- Nord, C. (1997). *Translating as a Purposeful Activity*. St. Jerome Publishing.
- Pérez-González, L. (2014). *Audiovisual Translation*. Routledge.
<https://doi.org/10.4324/9781315762975>
- Ren, D. (2023). Research on International Publicity Translation of Regional Cultural Tourism. *Foreign Language World*, No. 3, 32-39.
- Vermeer, H. J. (2004). Skopos and Commission in Translational Action. In A. Chesterman (Ed.), *Translation Quality Assessment* (pp. 190-211). St. Jerome.
- Wu, J., & Liu, Y. (2021). Multimodal Translation and National Image Construction Based on “Beautiful China” Promotional Films. *Journal of Sichuan International Studies University*, 33, 78-85.
- Wu, Y. (2021). Multimodal Translation Research Under the Medium Shift. *Foreign Languages*, 44, 115-120. (In Chinese)
- Yuan, X., & Zhu, C. (2025). Degree of Perception, Communication Context and Heterogeneous Components: An Examination of the Phenomenon of “Cultural Discount” in Film and Television Works in Cross-Cultural Communication. *Journal of Changsha University of Science and Technology (Social Sciences Edition)*, 40, 113-121. (In Chinese)
- Zhang, D. (2009). Exploration of a Comprehensive Theoretical Framework for Multimodal Discourse Analysis. *Chinese Foreign Languages*, 6, 24-30. (In Chinese)