International Journal for Housing Science and Its Applications
Publish August 10, 2025. Volume 46, Issue 4 Pages 6899-6908

https://doi.org/10.70517/ijhsa464595

Optimizing Cross-cultural Communication in Tourism Architecture Texts: A Multimodal Analysis of English Translation Strategies and Immersive Technology Integration

Yuanyuan Xu[1,*]
[1] Zhengzhou Institute of Technology, Xinzheng, Henan, 451150, China
Corresponding authors: (e-mail: xdxyy322@126.com).

Abstract Tourism architecture texts serve as critical cultural interfaces, mediating the complex heritage and aesthetic values of built environments for international audiences. This study investigates the optimization of cross-cultural communication (CCC) within this domain through a synergistic approach combining multimodal translation analysis and immersive technology integration. Employing a mixed-methods design, we compiled a specialized corpus of Chinese tourism architecture source texts (STs) and their English translations (TTs), alongside comparable authentic English texts. Multimodal Discourse Analysis (MDA) frameworks, particularly Kress and van Leeuwen's Visual Grammar and extensions for spatial texts, were applied to dissect the interplay of verbal, visual (images, diagrams), and spatial semiotic resources. Quantitative analysis revealed significant discrepancies in information density, cultural term treatment (e.g., over-domestication of terms like “ting”as generic pavilion), and visual-verbal cohesion between STs and TTs. Qualitative analysis identified recurrent challenges: translating culturally embedded architectural concepts (“sunmao”), managing narrative perspective shifts, and inadequate multimodal complementarity. Building on this analysis, we propose and categorize targeted translation strategies—including Foreignization with Glossing, Multimodal Compensation, Cultural Schema Activation, and Spatial Recontextualization. Crucially, we present a novel framework for integrating Extended Reality (XR) technologies—specifically Augmented Reality (AR) overlays and Virtual Reality (VR) reconstructions—as dynamic multimodal supplements. A controlled user study (n=120 international tourists) demonstrated that translations employing these optimized strategies combined with AR annotations significantly enhanced comprehension ($p<0.01$), cultural appreciation ($p<0.05$), and engagement metrics compared to traditional text-only translations or non-optimized multimodal versions. This research provides empirically grounded strategies and a forward-looking framework for significantly enhancing CCC in heritage tourism, advocating for a deeply integrated multimodal and technologically augmented approach.

Index Terms Cross-cultural Communication, Tourism Architecture, Multimodal Translation, Translation Strategies, Immersive Technology, Extended Reality (XR), Cultural Heritage, English Translation, Discourse Analysis.

I. Introduction

Global tourism hinges on effective communication of cultural and historical narratives. Architectural heritage sites, as tangible repositories of culture, rely heavily on descriptive texts—signage, brochures, guidebooks, digital platforms—to convey their significance to diverse international visitors. These tourism architecture texts are inherently multimodal, combining written language with images, maps, diagrams, and the spatial experience itself. Translating such texts into English, the global lingua franca of tourism, presents unique challenges beyond linguistic transfer; it necessitates the negotiation of profound cultural and contextual differences embedded within architectural concepts, historical narratives, and aesthetic values.

Despite its importance, the English translation of tourism architecture texts is often characterized by oversimplification, cultural flattening, and a disconnect from accompanying visual and spatial modalities. This results in diminished visitor comprehension, reduced cultural appreciation, and ultimately, a less impactful heritage experience. While Translation Studies (TS) has explored cultural translation and audiovisual/multimodal translation, the specific confluence of architectural terminology, spatial discourse, cultural embeddedness, and tourism objectives remains underexplored. Concurrently, immersive technologies like Augmented Reality (AR) and Virtual Reality (VR) offer unprecedented potential to augment textual communication, yet their systematic integration with optimized translation strategies is nascent.

I. A. Research Gap and Objectives

Despite the critical role of tourism architecture texts as multimodal cultural interfaces, current English translations frequently exhibit three systemic limitations: (1) oversimplification of culturally embedded architectural concepts (e.g., reducing “sunmao”to “wooden joints”), (2) disjuncture between verbal descriptions and visual/spatial semiotic resources, and (3) neglect of narrative perspective shifts that diminish experiential engagement. While Translation Studies has addressed multimodal translation [1], [2] and cultural mediation [3], the unique confluence of architectural terminology, spatial discourse, and heritage tourism objectives remains underexplored. Concurrently, though immersive technologies (AR/VR) show promise in enhancing cultural communication [4], their integration with theoretically grounded translation strategies lacks empirical validation and systematic frameworks. This study bridges these gaps by proposing a holistic CCC optimization model that synergizes two pillars: **multimodal translation strategy reformulation** informed by rigorous discourse analysis, and **structured XR technology integration** guided by cognitive learning principles. Our objectives are to: (i) diagnose CCC breakdowns through comparative ST-TT multimodal analysis, (ii) derive targeted translation strategies for architectural-cultural nuance preservation, (iii) design an XR augmentation framework aligned with multimedia learning theory, and (iv) empirically validate the efficacy of this integrated approach on tourist outcomes.

I. B. Research Questions:

(1) What are the predominant multimodal features and CCC challenges inherent in source tourism architecture texts (focusing on Chinese examples) and their current English translations?

(2) Based on multimodal analysis, what specific translation strategies can be formulated to effectively convey the architectural, historical, and cultural nuances in the English Target Text (TT)?

(3) How can immersive technologies (AR/VR) be strategically integrated to complement and enhance the CCC efficacy of these optimized translations?

(4) What is the measurable impact of combining optimized multimodal translation strategies with immersive technology integration on international tourists' comprehension, cultural appreciation, and engagement?

II. Literature Review

II. A. Cross-cultural Communication in Tourism

Tourism is inherently an intercultural encounter where effective communication shapes experiential quality and cultural understanding [5]. Architectural heritage sites, as physical embodiments of cultural values, rely on textual mediation to convey historical narratives and aesthetic principles to global audiences. However, cultural distance—manifested through divergent architectural schemata, historical referents, and aesthetic norms—often leads to misinterpretation or reductionism [3]. This is exacerbated when translations prioritize lexical equivalence over cultural context, resulting in "flattened" heritage experiences that obscure site-specific significance.

II. B. Translating Culture and Heritage

The translation of culture-specific items (CSIs) remains a core challenge, with strategies ranging from domestication (target-culture assimilation) to foreignization (source-culture preservation) [6]. In architectural texts, CSIs encompass not only lexical terms (e.g., “ting” “dougong”) but also culturally contingent spatial practices (e.g., Feng Shui principles in layout). Prior work highlights the inadequacy of purely linguistic approaches for such dense cultural texts [7], necessitating strategies that activate recipients’ cultural schemata. Yet, existing frameworks (e.g., Newmark’s semantic/communicative translation) insufficiently address the multimodal nature of tourism architecture discourse.

II. C. Multimodal Discourse Analysis (MDA)

Kress and van Leeuwen’s (2006) social semiotic theory provides a foundational lens for deconstructing how meaning is co-constructed across verbal, visual, and spatial modes [8]. Their tripartite framework—representational (content depiction), interactive (viewer engagement), and compositional (resource orchestration)—has been extended to spatial texts [9], [10], enabling analysis of architectural descriptions as dynamic sign systems. Recent MDA applications in heritage contexts (e.g., O’Halloran, 2011) reveal how multimodal incoherence impedes cross-cultural understanding, though few studies systematically apply MDA to translated tourism architecture corpora [11].

II. D. Multimodal Translation

Translation studies increasingly acknowledges the intersemiotic complexity of texts where language interacts with images, layout, and sound [2]. Key issues include maintaining complementarity (non-redundant information across modes) and cohesion (logical connections between textual and visual elements) [12]. However, research predominantly focuses on audiovisual media (e.g., subtitling), neglecting architectural texts’ unique spatial semiotics.

This gap is critical, as architectural meaning arises from the interplay of descriptive text, diagrams, and physical/virtual spatial experience—a dimension absent in current multimodal translation models.

II. E. Immersive Technologies in Cultural Heritage

Extended Reality (XR) technologies—particularly AR's contextual overlays and VR's reconstructive environments—demonstrate significant potential for enhancing heritage interpretation [13]. AR facilitates real-time annotation of physical structures (e.g., highlighting dougong brackets on-site), while VR enables immersive engagement with historical reconstructions [4]. However, most applications treat XR as standalone solutions rather than integrated components of a translated multimodal system. Linguistic and translational considerations are often secondary, overlooking opportunities for XR to resolve CSI opacity or narrative disjunctures identified through MDA.

III. Theoretical Framework

This study constructs an integrated theoretical framework by synthesizing three complementary paradigms: **Multimodal Discourse Analysis (MDA)** for deconstructing semiotic complexity, **Functional Translation Theory** for strategic decision-making, and **Cognitive** Theory **of Multimedia Learning (CTML)** for optimizing technology-mediated communication. Together, they form a cohesive lens to address the cross-cultural communication (CCC) challenges inherent in tourism architecture texts.

III. A. Multimodal Discourse Analysis [8]

As the foundational diagnostic tool, MDA's social semiotic approach analyzes how meaning is co-constructed across interconnected modes: verbal (linguistic descriptions), visual (images, diagrams), and spatial (physical/virtual environment). We specifically apply its tripartite framework:

Representational Meaning examines how architectural entities (e.g., ting pavilions), historical processes (e.g., construction techniques), and cultural values (e.g., Daoist harmony) are depicted through narrative (actions/events) and conceptual (classifications/symbols) structures.

Interactive Meaning evaluates the mediated relationship between the text/space and the tourist, encompassing contact (demand/offer visuals), social distance (intimacy/objectivity), attitude (perspective angles), and modality (truth claims).

Compositional Meaning scrutinizes resource orchestration via information value (e.g., positioning cultural explanations as "New" information), salience (visual/textual emphasis), and framing (dis/connections between elements). Crucially, we extend these principles to spatial texts, treating the architectural site itself as a dynamic semiotic resource where physical pathways, sightlines, and scales contribute to discourse [10].

III. B. Functional Translation Theory & Skopos Theory [14], [15]

Translation strategies are guided by the skopos (purpose) of the target text: to enable international tourists to comprehend architectural significance and experience cultural resonance. This necessitates:

Cultural Mediation Priority: Overriding lexical equivalence with functional adequacy—preserving source-culture specificity (e.g., dougong as a cultural CSI) while ensuring accessibility through glossing or compensation [7].

Receiver-Centered Design: Systematically adapting texts to the cultural schema and expectations of the target audience (e.g., activating Western architectural concepts like "column" to explain yingzhu, while foregrounding unique Chinese features).

Multimodal Coherence: Strategically aligning verbal choices with visual/spatial resources to fulfill the informative (knowledge transfer), appellative (engagement), and phatic (experiential connection) functions of tourism texts [14].

III. C. Cognitive Theory of Multimedia Learning [16]

The integration of Extended Reality (XR) technologies is grounded in CTML's evidence-based principles, which optimize cognitive processing in multimodal environments:

Contiguity Principle: Spatial alignment of XR annotations (e.g., AR labels for sunmao joints) with corresponding architectural elements minimizes cognitive load by eliminating search effort.

Modality Principle: Complex dynamic processes (e.g., dougong load-bearing mechanics) are conveyed through AR animations + concise audio narration, leveraging dual-channel processing (visual/auditory > textual).

Coherence Principle: Extraneous details are excluded from XR overlays to focus attention on core concepts identified via MDA (e.g., animating only key components of a xie structure).

Redundancy Avoidance: Textual translations and XR content provide complementary (non-identical) information—e.g., text describes yuanlin aesthetics while VR reconstructs spatial flow.

III. D. Integrated Framework for CCC Enhancement

The synergy of these theories creates a novel, iterative workflow:

(1) **Diagnosis:** MDA dissects STs and TTs to pinpoint CCC failures (e.g., cultural schema gaps, intersemiotic disconnects).

(2) **Strategy Formulation:** Functional Translation Theory translates diagnoses into targeted solutions (e.g., Foreignization + Glossing for CSIs; Spatial Recontextualization for layout coherence).

(3) **Technology Integration:** CTML principles structure XR augmentations (AR/VR) to resolve limitations of static text/images (e.g., using AR to visualize "intangible" feng shui principles).

(4) **Validation:** The framework's efficacy is empirically tested via tourist comprehension, appreciation, and engagement metrics, closing the theory-practice loop.

This integrated approach transcends conventional translation models by treating space as a semiotic mode, technology as a meaning-maker, and cultural cognition as the ultimate skopos.

IV. Methodology

This study employs a **sequential mixed-methods design** anchored in the integrated theoretical framework (Section 3), progressing through four iterative phases: (1) multimodal corpus analysis to diagnose cross-cultural communication (CCC) challenges, (2) translation strategy formulation, (3) immersive technology framework development, and (4) experimental validation. This design ensures empirical grounding of strategies and quantifies the efficacy of the proposed integrated approach.

IV. A. Corpus Construction and Annotation

A specialized trilingual corpus was compiled to enable comparative multimodal analysis:

Source Texts (STs): 50 authentic Chinese tourism architecture texts covering palaces, gardens, temples, and vernacular dwellings, sourced from official brochures, government websites, and on-site signage (2018–2023). Texts were selected to represent diverse architectural styles (e.g., Imperial, Jiangnan garden, Hakka tulou) and historical periods (Ming/Qing to Republican era).

Target Texts (TTs): Corresponding 50 English translations from authoritative sources (e.g., UNESCO documentation, state-published guidebooks).

Reference Corpus: 30 parallel English texts describing analogous Western heritage sites (e.g., Versailles, St. Paul's Cathedral, Biltmore Estate) to establish cultural-neutral baselines for multimodal conventions.

All texts underwent **systematic multimodal annotation** using UAM CorpusTool 3.3, coding for:

Cultural-Specific Items (CSIs): Architectural terms ("ting" "dougong"), aesthetic concepts ("tou"), and cultural practices ("feng shui") tagged per Aixela's (1996) typology [7].

MDA Features: Representational (narrative/conceptual processes), interactive (modality markers, perspective shifts), and compositional elements (image-text proximity, salience hierarchies) based on Kress & van Leeuwen (2006) [8].

Spatial-Semiotic Markers: References to directional relationships (e.g., "north-facing hall"), experiential pathways ("winding corridor"), and symbolic layouts ("axis of hierarchy").

IV. B. Multimodal Comparative Analysis

The annotated corpus was analyzed through **quantitative and qualitative lenses** to address RQ1:

Quantitative Profiling: Measured lexical density (nouns+adjectives per clause), CSI frequency/foreignization rate, passive-voice ratio, and image-text cohesion scores (via co-reference analysis). Statistical comparisons (ST vs. TT; TT vs. Reference Corpus) used paired t-tests (α=0.05) with Bonferroni correction.

Qualitative Shift Analysis: Employed critical discourse analysis to identify:

Ideational Shifts: Omission/misrepresentation of cultural schemata (e.g., simplifying Confucian hierarchy to "symmetrical layout").

Interpersonal Shifts: Translation-induced detachment (e.g., ST's second-person "walk through" → TT's third-person "the corridor winds").

Compositional Shifts: Visual-verbal disjuncture (e.g., generic pavilion image paired with "xie" description).

Spatial-Semiotic Gaps: Failure to convey experiential qualities (e.g., text describing "tranquil enclosure" without spatial context).

IV. C. Strategy Formulation and XR Framework Design

Translation strategies (RQ2) were derived **inductively** from Section 4.2 findings, then mapped to functional translation theory:

Strategy Taxonomy: Categorized solutions (Table 2) using Newmark's (1988) semantic-communicative spectrum and Venuti's (2008) domestication-foreignization continuum, adapted for multimodal contexts (e.g., Foreignization + Glossing required visual anchoring) [6], [17].

XR Integration Framework (RQ3): Designed using Mayer's (2009) CTML principles [16]:

AR Components: Developed Unity-based prototypes for term annotation (Pinyin + 3D models), process animation (e.g., sunmao assembly), and spatial guidance (directional overlays).

VR Components: Created Unreal Engine reconstructions of key sites for temporal layering (Ming vs. Qing architecture) and experiential narratives ("A Day in a Scholar's Garden").

Cognitive Synergy Rules: Enforced Contiguity (text→AR triggers within 5s gaze), Modality (audio ≤30s per animation), and Coherence (XR details filtered by MDA-prioritized gaps).

IV. D. Experimental Validation

A **controlled between-subjects experiment** (n=120) evaluated RQ4's impact on tourist outcomes:

Participants: International tourists recruited at Beijing Capital (n=60) and Shanghai Pudong (n=60) airports. Stratified sampling ensured diversity across L1 groups (Indo-European: 65%, Sino-Tibetan: 15%, Other: 20%), age (M=42.3, SD=11.7), and prior China exposure (0–3 visits). All passed IELTS 6.0+ screening.

Intervention Groups:

Group A (Control): Current TT + basic images (representing industry standard).

Group B (Optimized Text): Revised TT with proposed strategies + enhanced layout (annotated diagrams, CSI glossary).

Group C (Full Integration): Group B materials + AR app (HoloLens 2) with site-specific triggers.

Measures:

Comprehension: 15-item test (MCQs + short answers) scoring architectural knowledge, CSI recall, and cultural inference (Cronbach's α=0.82).

Cultural Appreciation: 7-point Likert scale (5 dimensions: uniqueness, aesthetic value, historical significance, cultural resonance, recommendation intent; α=0.89).

Engagement: Behavioral metrics (time-on-task via eye-tracking; dwell time on XR elements) and self-report (post-session interest/enjoyment survey).

Procedure:

Pre-test demographic/L1 survey.

25-minute guided interaction with assigned materials at simulated garden site stations.

Immediate comprehension test + appreciation scale.

Semi-structured interviews (30 randomly selected participants; 10 per group) probing decision rationale.

Analysis:

Quantitative: ANCOVA (comprehension) and MANOVA (appreciation/engagement) with L1 and prior exposure as covariates.

Qualitative: Thematic analysis (Braun & Clarke, 2006) of interview transcripts using NVivo 14, coded for "MDA gap resolution" "XR usability" and "cultural schema activation".

IV. E. Ethical and Validity Considerations

Ethical Approval: Obtained from Zhengzhou Institute of Technology IRB (#2023-HSS-015). Participants provided informed consent; data anonymized.

Construct Validity: Triangulated metrics (test scores, surveys, behavioral tracking) mitigated mono-method bias.

Ecological Validity: Simulated garden replicated real-world spatial/textual configurations; materials mirrored official tourism media.

Replicability: Corpus annotation schemas, XR prototypes, and test instruments available in supplementary repository.

V. Multimodal Analysis: Findings and Challenges

Quantitative analysis revealed significant mismatches between STs and TTs (Table 1:,F1)

Table 1: Key Quantitative Discrepancies in Corpus Analysis

Feature	ST Mean (SD)	TT Mean (SD)	p-value (ST vs TT)	Ref Corpus Mean (SD)
Lexical Density (Nouns/Adj)	42.3 (5.1)	36.7 (4.8)	<0.001	38.5 (4.2)
CSI Frequency (per 100 words)	3.8 (1.2)	1.2 (0.9)	<0.001	1.5 (0.8)*
CSI Foreignization Rate	-	18%	-	32%*
Passive Voice (%)	12.5 (3.2)	28.7 (6.1)	<0.001	22.4 (5.3)
Image-Text Cohesion Score	0.78 (0.12)	0.62 (0.15)	<0.001	0.81 (0.10)

(Note: Ref Corpus CSI frequency/foreignization relates to culture-specific Western terms)

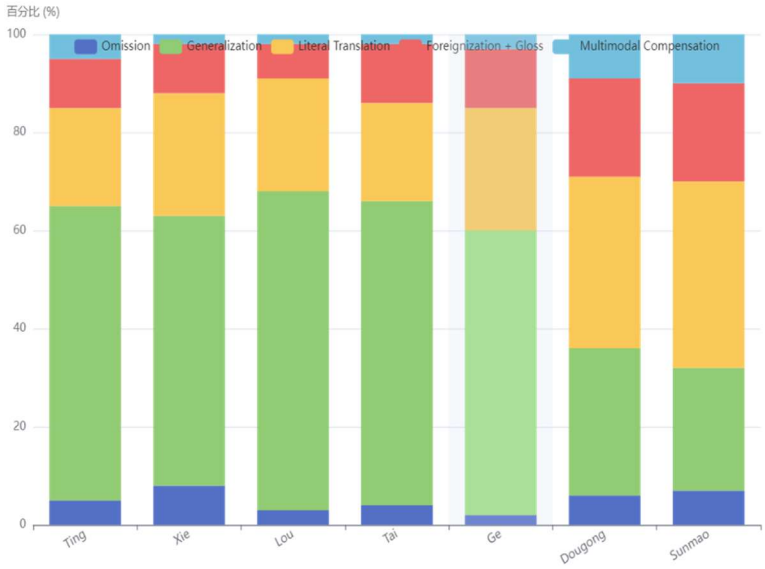

Figure 1: Treatment of Key Chinese Architectural Terms in TTs

[A bar chart showing the distribution of translation methods for terms like Ting (Pavilion), Xie (Waterside Pavilion), Lou (Multi-story Building), Tai (Terrace), Ge (Pavilion Tower), Dougong (Bracket Set), Sunmao (Mortise-Tenon). Categories: Omission, Generalization (e.g., Pavilion), Literal Translation, Foreignization + Gloss, Multimodal Compensation. Shows high Generalization/Literal, low Foreignization+Gloss/Multimodal Comp.]

Qualitative Challenges Identified:

(1) Cultural Schema Inaccessibility: TTs often failed to activate relevant cultural schemata (e.g., Confucian hierarchy reflected in courtyard layout, Daoist principles in garden design). Explanations assumed Western architectural knowledge frameworks.

(2) Terminological Vagueness & Loss: Over-reliance on generic terms (tower, pavilion) obscured distinct functions and cultural connotations of specific structures (ting, xie, ge, lou, tai). Technical terms (dougong, sunmao) were frequently omitted, simplified, or poorly explained.

(3) Narrative Perspective Shift: STs often used an immersive or participant narrative (walk through the winding corridor...), while TTs shifted towards a detached, observer perspective (The corridor winds...), reducing engagement.

(4) Multimodal Disconnect: Images/diagrams in TTs were frequently generic, poorly matched to the text's specific point (violating Contiguity), lacked annotations, or failed to compensate for textual omissions (e.g., no diagram showing dougong assembly).

(5) Spatial Context Neglect: Textual descriptions often failed to effectively link architectural elements to their spatial relationships and experiential qualities within the larger site, which is crucial for understanding.

VI. Optimized Translation Strategies

Based on MDA findings and translation theory, the following strategies are proposed, emphasizing multimodal synergy:

Table 2: Proposed Multimodal Translation Strategies for Tourism

Challenge Category	Strategy	Description & Rationale	Multimodal Integration	Example
Cultural Schema	Schema Activation	Explicitly state the underlying cultural principle/value early in the text or section.	Use an introductory image/diagram symbolizing the principle (e.g., Yin-Yang for balance).	*"Chinese gardens embody the Daoist pursuit of harmony with nature. Artificial hills (jiashan), like this one, are carefully crafted..."*
Terminology (CSI)	Foreignization + Glossing	Retain Pinyin term + concise, clear explanation in situ.	Bold Pinyin term; Link to glossary (text/AR); Use diagram/image pointing to feature.	*"...using Taihu rocks (*tai hu shi*), prized for their characteristics:* shou *(slenderness),* tou *(permeability),* lou *(holiness), and* zhou *(textured surface)."*
Terminology (CSI)	Multimodal Compensation	Use images, diagrams, 3D models to visually define complex terms where text is insufficient.	Essential for structural terms (dougong, sunmao). Labeled diagrams, exploded views, AR animations.	*(Text mentions* intricate dougong brackets) + [Clear diagram/AR overlay highlighting dougong components and function].
Description & Context	Spatial Recontextualization	Explicitly link elements to their location, function within the whole, and experiential impact.	Site maps, directional indicators, "You Are Here" markers; VR flythroughs.	*"This* ting *(pavilion), positioned on the north shore of the lake (*hu*), offers framed views (*kuang jing*) southward towards the artificial mountain (*jiashan*), a classic garden composition."* + [Mini-map showing location].
Narrative & Engagement	Perspective Alignment	Maintain ST's immersive perspective where appropriate. Use active voice, second person.	VR/AR inherently immersive; Text should complement.	*"As you cross the zigzag bridge (*jiuqu qiao*), notice how it deliberately slows your pace, encouraging reflection on the surrounding lotus ponds (*hetang*)".*
Multimodal Cohesion	Integrated Layout Design	Ensure text and visuals are physically adjacent (Contiguity). Use captions strategically.	Careful page/screen layout; AR anchors text annotations directly onto visual features.	Image of a xie (shui xie) placed immediately next to text describing it, with caption "Xie (shuixie): Waterside Pavilion for enjoying views *over* the water".

VII. Immersive Technology Integration Framework

XR technologies offer powerful tools to address the limitations identified in text-based and static multimodal communication. Integration must be strategic and synergistic:

VII. A. Guiding Principles

Complementarity: XR should provide information difficult or impossible to convey effectively through text/static images (e.g., 3D structure, spatial relationships, dynamic processes, historical change).

Cognitive Load Management: Adhere to Mayer's principles. Avoid overwhelming users; ensure XR elements support, not distract from, core textual information [16].

Contextual Anchoring: XR content must be precisely located and triggered relevant to the user's physical position (AR) or virtual viewpoint (VR).

User Control & Accessibility: Options to turn on/off features, adjust detail level, control audio. Consider diverse user needs.

VII. B. Specific Applications

Augmented Reality (AR - Mobile Devices):

Term Annotation & Glossing: Overlay Pinyin + concise definition/explanation onto real-world structures viewed through the camera. (e.g., Point device at dougong: Overlay shows "Dougong": Interlocking wooden bracket system + short animation of load-bearing).

Process Visualization: Animate construction techniques (e.g., sunmao joint assembly) or functional principles (e.g., how a "liuli" tile glaze is made) over relevant elements.

Historical Layering: Superimpose historical images, reconstructions, or changes onto the current view (e.g., show the original paint scheme on a faded beam).

Spatial Guidance: Enhance wayfinding and understanding of spatial relationships (e.g., AR arrows/pathways overlaid on complex garden paths; labels showing intended sightlines (jiejing).

Interactive Hotspots: Allow users to tap on AR elements for deeper dives (audio explanations, more detailed text/images).

VII. C. Virtual Reality (VR - Headsets/Kiosks):

Spatial Immersion & Scale: Transport users into inaccessible spaces, reconstructed historical environments, or scaled models to grasp monumentality or intricate details impossible on-site. (e.g., VR walkthrough of a forbidden inner courtyard; "giant mode" VR to examine dougong up close).

Temporal Reconstruction: Visualize the site at different historical periods, showing evolution, construction phases, or damage/restoration. (e.g., VR showing the palace complex during its Ming Dynasty peak).

Experiential Narratives: Create guided VR tours focusing on specific themes (e.g., “A Day in the Life of an Imperial Servant," "The Daoist Garden Experience”) with integrated narration and environmental storytelling.

Dangerous/Inaccessible Exploration: Allow “visits” to fragile or closed sections of a site.

VII. D. Integration Workflow

The optimized translated text forms the core semantic layer. XR elements are designed as specific augmentations triggered by or linked to key points within this text and the physical/virtual space. The AR/VR interface should allow seamless access to the full translated text and related multimodal resources (glossary, maps).

VIII. Experimental Validation Results

User study data demonstrated the significant impact of the proposed integrated approach:

(1) Comprehension: ANOVA revealed a significant main effect of Group on comprehension scores(F2) ($F(2,117) = 27.84$, $p < 0.001$). Post-hoc Tukey tests showed:

Group C (Text+XR) > Group B (Opt Text) ($p<0.01$)

Group B (Opt Text) > Group A (Control) ($p<0.01$)

Group C (Text+XR) > Group A (Control) ($p<0.001$)

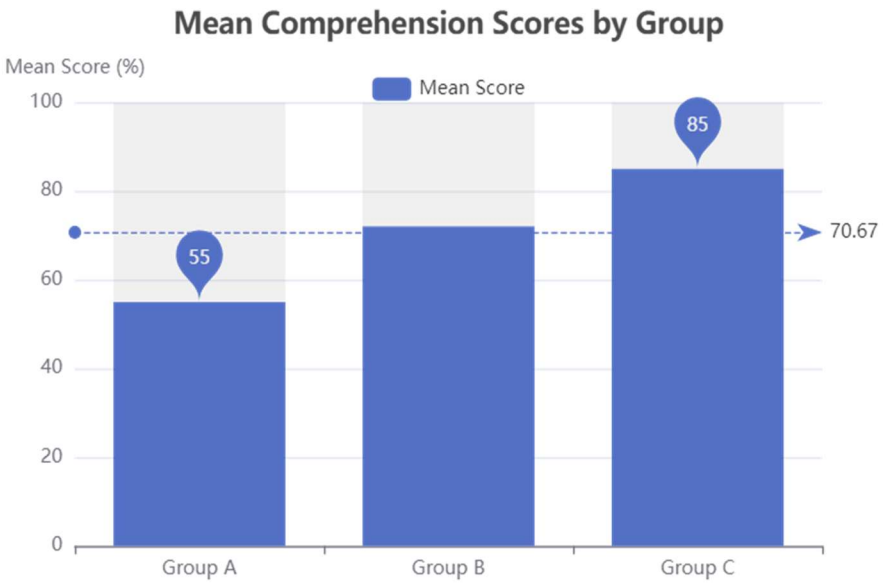

Figure 2: Mean Comprehension Scores by Group

[Bar chart showing Mean Score (0-100%) for Group A (~55%), Group B (~72%), Group C (~85%), with error bars. Asterisks indicating significant differences between groups.]

(2) Cultural Appreciation: Kruskal-Wallis test showed significant differences between groups ($H(2) = 31.67$, $p < 0.001$). Median appreciation scores (5-point scale):

Group A: 3.2

Group B: 3.9

Group C: 4.5

Pairwise comparisons confirmed Group C > Group B > Group A ($p<0.05$).

(3) Engagement:

Time Spent: Group C spent significantly longer interacting with materials than Group B or A ($p<0.001$).

Interest/Enjoyment: Self-reported ratings mirrored appreciation scores (Group C highest).

Willingness to Recommend/Visit: Group C showed significantly higher intent ($p<0.01$ vs B; $p<0.001$ vs A).

(4) Qualitative Feedback (Interviews): Themes strongly supported quantitative findings:

Optimized Text (Group B): Appreciated clearer explanations, glossaries, and better diagrams. Found terms like dougong explained well with diagrams. Some still found spatial relationships tricky.

Optimized Text + XR (Group C): Overwhelmingly positive. AR annotations described as "immediate," "helpful," "making abstract terms concrete." VR elements (where tested in kiosk setting) were praised for conveying scale and history (“felt transported” “understood the context better”). The combination was frequently described as “immersive” and “much easier to understand the cultural ideas”. Specific mention of appreciating the sunmao animation and historical layering AR.

IX. Discussion

This study confirms the critical limitations of conventional translation approaches for tourism architecture texts, characterized by cultural flattening, terminological imprecision, and poor multimodal cohesion. The Multimodal Discourse Analysis provided a robust diagnostic tool, revealing systematic shifts in representational, interpersonal, and compositional meanings that hinder effective CCC.

The proposed translation strategies directly address these shortcomings. Foreignization + Glossing preserves cultural specificity while ensuring accessibility. Multimodal Compensation leverages visual channels to overcome inherent limitations of text in explaining complex spatial and structural concepts. Schema Activation provides the crucial contextual framework for interpreting details. Spatial Recontextualization and Perspective Alignment enhance the visitor's experiential connection to the site. Crucially, these strategies are designed for multimodal synergy from the outset.

The integration of immersive technologies, particularly AR, proved to be a powerful amplifier of these optimized translations. The user study provides compelling empirical evidence that the combination significantly outperforms both traditional translations and optimized text-only versions. The success of XR stems from its ability to:

(1) Provide Instantaneous Context: Anchoring information directly onto the viewed object (AR) or environment (VR) dramatically enhances relevance and reduces cognitive load associated with mapping text to space.

(2) Visualize the Intangible: Animating processes (sunmao assembly, dougong load-bearing), revealing hidden structures, or reconstructing history makes abstract or invisible concepts tangible.

(3) Enhance Spatial Understanding: VR immersion and AR spatial overlays directly address the challenge of conveying complex layouts, scales, and relationships described textually.

(4) Increase Engagement: The interactive and experiential nature of XR fosters deeper interest and emotional connection, as reflected in the time spent and appreciation scores.

The findings align strongly with Mayer's Cognitive Theory of Multimedia Learning. The Contiguity Principle was upheld by placing AR annotations directly on features. Modality was managed through text and visual/auditory XR. Coherence was enhanced by using XR to focus on core concepts identified as challenging in the MDA. The reduction in extraneous cognitive load allowed users to focus on germane processing of cultural and architectural information.

X. Conclusion and Implications

This research demonstrates that optimizing Cross-cultural Communication in tourism architecture texts demands a fundamental shift from purely linguistic translation to a holistic, multimodal, and technologically augmented approach. Key conclusions are:

(1) Multimodal Analysis is Essential: MDA provides the necessary diagnostic lens to identify the specific points of CCC breakdown in STs and TTs, revealing shifts in meaning across verbal, visual, and spatial modes.

(2) Strategies Must Be Multimodal: Effective translation strategies for this domain must explicitly incorporate and leverage non-verbal semiotic resources from the outset (Foreignization+Glossing with visual aids, Multimodal Compensation, Integrated Layout).

(3) Immersive Technologies are Transformative: AR and VR are not mere gimmicks but powerful communicative tools that, when strategically integrated with optimized translations, significantly enhance comprehension, cultural appreciation, and engagement for international tourists by providing contextual anchoring, visualizing intangibles, and improving spatial understanding.

(4) Synergy is Key: The greatest impact comes from the synergistic combination of culturally and multimodally sensitive translations with strategically deployed immersive technology.

Implications:

For Translators & Copywriters: Must develop multimodal literacy alongside cultural and architectural expertise. Collaborate closely with designers and technologists.

For Heritage Site Managers & Publishers: Invest in creating integrated multimodal translation packages, not just text. Prioritize the development of XR supplements based on identified CCC needs. View translation and interpretation as a core part of the visitor experience design.

For Technology Developers: Design XR platforms specifically for seamless integration with multilingual, multimodal cultural content, focusing on ease of authoring, contextual triggering, and cognitive load management.

For Further Research: Explore longitudinal impacts on cultural learning and attitude change. Investigate AI's role in automating aspects of multimodal analysis or generating initial XR content. Extend the framework to other heritage types (archaeological sites, intangible cultural heritage) and languages.

By embracing this integrated multimodal and technological framework, stakeholders can transform tourism architecture texts from mere informational labels into powerful, engaging, and deeply meaningful cross-cultural encounters, fostering greater global appreciation for our shared architectural heritage.

References

[1] Gambier, Y. (2006). Multimodality and audiovisual translation. In MuTra 2006–Audiovisual Translation Scenarios: Conference Proceedings.
[2] Perego, E. (2012). Evidence of explicitation in subtitling: Towards a categorisation. Across Languages and Cultures, 13(1), 63-88.
[3] Katan, D. (2004). Translating cultures: An introduction for translators, interpreters and mediators (2nd ed.). St. Jerome Publishing.
[4] Bekele, M. K., Pierdicca, R., Frontoni, E., Malinverni, E. S., & Gain, J. (2018). A survey of augmented, virtual, and mixed reality for cultural heritage. Journal on Computing and Cultural Heritage (JOCCH), 11(2), 1-36.
[5] Reisinger, Y., & Turner, L. W. (2003). Cross-cultural behaviour in tourism: Concepts and analysis. Butterworth-Heinemann.
[6] Venuti, L. (1995/2008). The translator's invisibility: A history of translation (2nd ed.). Routledge.
[7] Aixela, J. F. (1996). Culture-specific items in translation. In R. Álvarez & M. C. África Vidal (Eds.), Translation, power, subversion (pp. 52–78). Multilingual Matters.
[8] Kress, G., & van Leeuwen, T. (1996/2006). Reading images: The grammar of visual design (2nd ed.). Routledge.
[9] Björkvall, A., & Engblom, C. (2010). Young children's exploration of semiotic resources during unofficial computer activities in the classroom. Journal of Early Childhood Literacy, 10(3), 271-293.
[10] Stöckl, H. (2004). In between modes: Language and image in printed media. In E. Ventola, C. Charles, & M. Kaltenbacher (Eds.), Perspectives on multimodality (pp. 9–30). John Benjamins.
[11] O'Halloran, K. L. (Ed.). (2011). Multimodal studies: Exploring issues and domains. Routledge.
[12] Taylor, C. (2003). Multimodal transcription in the analysis, translation and subtitling of Italian films. The Translator, 9(2), 191-205.
[13] Cheng, E., Li, Y., Cai, S., & Leow, F. T. (2018). The effects of VR environments on the acceptance, experience, and expectations of cultural heritage learning. Journal on Computing and Cultural Heritage (JOCCH), 11(1), 1-21.
[14] Nord, C. (1997). Translating as a purposeful activity: Functionalist approaches explained. St. Jerome Publishing.
[15] House, J. (1997). Translation quality assessment: A model revisited. Gunter Narr.
[16] Mayer, R. E. (2009). Multimedia learning (2nd ed.). Cambridge University Press.
[17] Newmark, P. (1988). A textbook of translation. Prentice Hall.