

Multimodal Approaches in Teaching Chinese as a Foreign Language: A Study of Idioms in TCFL Microlessons

Shuo Cao¹, Zongbo He², Siyu Wang¹, and Yiqi Xu^{1*}

¹*School of Foreign Language, Dalian University of Technology, Dalian, China*

²*Dalian Hi-Tech Industrial Park Second School, Dalian, China*

**Corresponding author: Yiqi Xu.*

Abstract

Microlessons have emerged as an effective means for foreign learners to acquire Chinese in technology-enhanced education. As a video-based format, microlessons typically integrate multiple semiotic resources, including language, images, sound, and gestures. However, in the context of Teaching Chinese as a Foreign Language (TCFL), particularly in idiom instruction where metaphorical meaning and cultural load are prominent, how multimodal resources function together across different instructional stages remains less explored. Therefore, the present study adopts a mixed-methods design combining qualitative and quantitative approaches to examine patterns of multimodal use and their pedagogical effects in two award-winning TCFL idiom microlesson series. The results show that teachers' speech, on-screen text, and images constitute the core modalities, with auditory–visual complementarity as the dominant pattern, while kinesthetic cues play a reinforcing role at key explanatory moments. Moreover, modality choices also vary across instructional stages and content types. Based on these findings, the study recommends complementary, functionally appropriate modality design guided by adaptability, efficiency, and economy to improve metaphor comprehension and retention.

Keywords

microlessons, teaching Chinese as a foreign language, multimodal teaching, idioms instruction

1. Introduction

Microlessons have become an increasingly common format in technology-enhanced education for their short duration, focused content, and flexible access across online learning [1, 2]. Early definitions typically describe a microlesson as a brief instructional video that focuses on a single learning point within a limited time span [3]. In China, the concept of microlesson was first systematically introduced as a video-centered integration of teaching resources organized around a specific knowledge point or a particular instructional stage [4]. In this sense, microlessons offer an efficient way to present focused content efficiently, allowing learners to access key points quickly and review them when needed.

In language education, these features make microlessons especially suitable for learning beyond the classroom, as learners can revisit the lessons at their own pace and reinforce key points after instruction [5].

This is particularly relevant in Teaching Chinese as a Foreign Language (TCFL), where learning often involves cross-cultural communication and culturally embedded expressions that may not be fully grasped for the first time. Therefore, TCFL learners can benefit from additional opportunities for preview and review through microlessons, which support comprehension and help consolidate learning over time.

Within this TCFL context, Chinese idioms represent a high-value yet demanding domain because idiom interpretation is often culturally grounded and metaphorically embedded. Learners' comprehension therefore depends not only on lexical meaning but also on linking linguistic forms to culturally shared conceptual associations and prior knowledge [6]. Since idiom microlessons are video-based and multimodal, they can realize culturally grounded metaphorical mappings through the coordinated use of speech, on-screen text, images, and sound [7], thereby supporting learners' idiom comprehension. More specifically, visual resources can provide concrete scenes that activate source-domain knowledge, while teachers' speech and on-screen text can guide learners toward target-domain interpretation. Overall, idiom microlessons provide an effective multimodal pathway for making culturally embedded metaphorical meanings more accessible to TCFL learners.

However, existing research on microlessons has largely focused on learners' willingness to use microlessons and their learning outcomes, often relying on self-report measures and technology-acceptance perspectives to explain adoption intentions and perceived effectiveness [5, 8, 9]. Despite these contributions, these studies are largely outcome-oriented, while how to enhance microlesson quality through the design and stage by stage coordination of multimodal resources remains underexplored.

To address this gap, the present study analyzes two award-winning TCFL idiom microlesson series through multimodal discourse analysis and triangulates the qualitative patterns with learner and expert evaluations, aiming to identify the multimodal strategies across instructional stages that facilitate metaphor-based idiom comprehension.

1 Literature Review Microlessons and Multimodal Instruction in TCFL

In recent years, the scope of microlesson research typically falls into three main aspects. The first concerns microlesson design, focusing on how content, teaching stages, and learner groups are aligned and organized within a short instructional unit [4, 10]. The second examines pedagogical integration with broader instructional models, which explores how they complement other classroom practices [11, 12]. The third addresses effectiveness and evaluation, assessing learning outcomes and learner experience through measures such as tests, questionnaires, and interviews [5, 13, 14]. However, many of these studies have been conducted in general education or first-language instructional settings, whereas research in TCFL remains comparatively limited.

Importantly, microlessons are also multimodal by nature, typically combining speech, on-screen text, images, and sound within tightly timed instructional stages. In TCFL, where learning often involves cross-cultural communication and culturally embedded meanings, understanding how these semiotic resources are coordinated across stages is therefore pedagogically important. However, research on TCFL microlessons has been largely practice-oriented. Many studies take prize-winning microlesson competition videos as cases, summarize common design features, and offer suggestions for improving how microlessons are produced and used in teaching. For example, Li analyzed prize-winning TCFL microlessons on near-synonym discrimination and summarized how target points are presented, sequenced, and practiced, while similar case-based work on culture-themed microlessons discusses how cultural content is condensed, visualized, and made interactive within short video units [15, 16]. However, most of this work stays at the level of pedagogical design guidance for microlessons.

A small but growing body of research has begun to examine TCFL microlessons from a multimodal perspective, describing how semiotic resources are used across instructional stages [17,18]. Yet the evidence is still largely descriptive, offering limited guidance on which modality combinations are most appropriate for particular teaching stages or how multimodal choices relate to learning outcomes. To address this gap, this study adopts a multimodal coordination perspective to examine modal coordination across teaching stages and their pedagogical functions in idiom microlessons.

1.2 Theoretical Foundations of Multimodal Coordination

Multimodality has been defined as both a theoretical and applied perspective on how meaning is made through the integration of language, visual resources, and other semiotic systems in traditional and digital media [19]. In educational settings, multimodal resources such as images and animation increasingly play a central role in meaning communication, which makes instructional design inseparable from how modes are selected and coordinated [20].

In the field of TCFL, research on multimodality primarily focuses on its application in teaching, analysis of multimodal functions, and studying multimodal interaction and synergy [7, 21-24]. Across these studies, a shared point is that modes rarely work separately in classrooms. Teachers typically combine speech, on-screen texts, visuals, and body language to make meanings clearer. As a result, understanding how these modes are coordinated in instruction is of particular importance, since effective teaching in multimodal settings relies on selecting and sequencing modes in ways that fit the demands of different teaching stages.

To examine modal coordination systematically, the present study adopts Zhang's framework of multimodal relations [7]. Zhang argues that a single mode often cannot fully convey meaning, which is why additional modes are introduced to supplement interpretations [7]. The framework distinguishes complementary and non-complementary relations and specifies relation types such as highlighting, prioritizing, expanding, coordinating, embedding, and overlapping. Table 1 summarizes these definitions and is used as the theoretical framework in the present study.

Table 1: Definition of Multimodal Relations

Type	Relation of modalities	Interrelation	Definition
Complementary	Reinforced	Highlighting	Modality 1 serves as a background and highlights modality 2.
		Prioritizing	Modality 1 is the primary means of communication, while modality 2 assumes a secondary role and provides support.
		Expanding	Modality 1 complements what modality 2 does not adequately express.
	Non-reinforced	Coordinating	Modality 1 and 2 collaborate in conveying a complete meaning.
		Combining	In the same modality, meaning is shared through different types of media channels.
Intersecting		Modality 1 and 2 are used interchangeably to indicate either a specific time or an activity.	
Non-complementary	Embedding	Part-whole	Modality 1 conveys the complete semantic content, while modality 2 only replicates a portion of the information.
		Abstract - concrete	Modality 1 represents an abstract summary, while modality 2 provides concrete examples.
	Overlapping	Redundant	Modalities 1 and 2 deliver identical information with no additional or reduced meaning—used for repetition or convenience.
		Resistant	Modality 1 and 2 conflict or negate each other, conveying opposite or contradictory meanings
		Exclusive	Modality 1 and 2 are mutually exclusive, valid only in their respective contexts and not co-present.
	Contextual Interaction	Independent	Each modality can be interpreted separately without relying on others.
		Dependent	Modality 1 must rely on modality 2 for interpretation; meaning arises from contextual linkage

In addition to relation types, Zhang proposes principles for modality collocation that are widely used in multimodal teaching research in China [7]. The principles of multimodal collocation mainly include the principles of effectiveness, adaptation, and choice. The effectiveness principle emphasizes that the use of instructional modalities should serve learning objectives. The adaptation principle highlights that different modes should be integrated in a way that fits the instructional sequence and works together coherently. The economy principle concerns using modes efficiently and avoiding unnecessary complexity. If two modes are sufficient to achieve the intended instructional effect, introducing the third mode is likely to be unnecessary.

Effective multimodal design involves finding a balance between using enough modes to support learning and avoiding unnecessary complexity.

Although multimodal teaching has been widely discussed, two limitations remain when it comes to TCFL microlessons. First, compared with the extensive studies on multimodal teaching, relatively few have examined multimodality in the specific context of TCFL, especially in microlesson settings. Second, the limited work that does address TCFL microlessons tends to describe modal distribution but provides less analysis of how modes relate to one another and how modalities align with different microlesson stages. Therefore, this study builds on multimodal discourse analysis framework, with particular attention to intermodal synergic relations framework to examine how auditory, visual, and kinesthetic resources are coordinated across teaching stages in TCFL microlessons and to provide implications for microlesson design [7].

1.3 Teaching Chinese Idioms in TCFL Through Multimodal Instruction

Chinese idiom instruction is widely recognized as an important component of TCFL as idioms often express cultural meanings that are not clear from the literal words [25]. At the same time, idioms are difficult for L2 learners because understanding them often requires cultural background knowledge and knowing how they are used in context. These difficulties also mean that idiom teaching is not only about giving a definition. Teachers usually need to help learners build a clear meaning, connect the idiom with a cultural story or image, and learn how the idiom works in real contexts, such as typical collocations, register, and pragmatic restrictions. In practice, TCFL idiom instruction often emphasizes origin explanation, paraphrasing, and example sentences. For instance, Wu and Wang analyzed how example sentences are used in idiom teaching and stressed their role in presenting usage patterns [26], while Wang and Luo highlighted the importance of cultural explanation and contextualized input [25]. However, these approaches can be less effective when time is limited and when learners have little cultural knowledge to rely on.

From this perspective, multimodal instruction is especially applicable for idiom learning. Visual resources like pictures and animation can make the idiom image and story easier to understand, while spoken explanation and on-screen text can guide attention to key meaning components. In some cases, embodied resources such as gesture can further strengthen memory by linking language to action. For microlessons, multimodality is a way to make short instruction more efficient by showing key images, highlighting meaning parts, and reducing learners' guessing load within a limited time window. Evidence from language learning research also suggests that learning outcomes can vary with different modalities, which further supports the need to design and coordinate modes carefully in video-based instruction [27].

A useful theoretical basis for this is embodied cognition, which views meaning as grounded in body experiences. In language learning, gestures can provide an embodied link between form and meaning, and studies have shown that words learned with gestures are often remembered better than words learned through verbal input alone [28, 29]. This embodied account is relevant to idioms because many idioms are image-based and require learners to move beyond literal words to a figurative interpretation. In cognitive terms, idiom processing typically involves both literal meaning activation and figurative interpretation and can be analyzed using mechanisms that are widely discussed in metaphor processing [30]. In fact, research in figurative language suggests that conceptual metaphor can play an active role in idiom comprehension, indicating a cognitive connection between idioms and metaphor processing [31]. In microlessons, multimodal resources can therefore provide a visual representation of source-domain information, which supports the construction of metaphor-based idiom meaning. Despite these theoretical insights, little research has examined how metaphor-based idiom meaning is constructed through the coordinated use of multimodal resources in TCFL idiom microlessons, especially in relation to different instructional stages and their pedagogical functions.

To address the above gaps, the present study seeks to answer the following research questions:

- (1) What multimodal resources are used in TCFL idiom microlessons, and what is their overall distribution and interaction pattern?
- (2) How are auditory, visual, and kinesthetic modalities coordinated across different instructional stages in TCFL idiom microlessons?
- (3) How is metaphor-based idiom meaning constructed through multimodal resources in TCFL idiom microlessons?

(4) What implications can be drawn for the multimodal design of TCFL idiom microlessons?

2. Method

2.1 Research Design

This study adopted a mixed-methods design to examine multimodal distribution, relational patterns, and metaphorical meaning construction in two award-winning TCFL idiom microlesson series. For the qualitative analysis, ELAN-based annotation and multimodal relation coding were conducted across instructional stages. The quantitative analysis involved descriptive statistical analyses of the annotated multimodal data and evaluative questionnaire results, which were used to identify overall patterns of modal use and to triangulate the qualitative findings with learner and expert feedback.

2.2 Materials

This study selected six first-prize-winning idiom microlesson videos from the second and third sessions of the Tangfeng Chinese Microlesson Competition as research materials [15]. The dataset consists of two series (Series I and Series II), with three videos in each series.

The Tangfeng competition is a well-established national platform in international Chinese language education, and its award-winning videos provide a relatively stable benchmark for examining multimodal teaching design. Idiom-focused microlessons are particularly suitable for multimodal analysis because idiom meaning is often metaphorical and culturally grounded.

Two sets of winning works were included and labeled as Series I and Series II. Series I (third session) targets intermediate-to-advanced learners from diverse L1 backgrounds and focuses on idioms related to zodiac culture (HSK 4 and above). The three videos last 9:42, 9:56, and 9:47. Series II (second session) targets native English-speaking university learners at HSK 4 and above (approximately 2,500 common words). The three videos last 9:53, 9:23, and 9:35.

The selection of these two series followed three criteria. First, only first-prize-winning works were included to ensure high instructional quality and pedagogical validity. Second, all selected videos focused on idiom instruction, allowing for direct comparison of multimodal strategies in metaphor-rich content. Third, the two series differ in learner profile and instructional design, which enables comparative analysis of multimodal choices across learner groups while maintaining overall comparability in lesson length and teaching objectives.

By analyzing these materials, the study aims to identify shared multimodal characteristics of effective TCFL idiom microlessons and to explore how different multimodal configurations support metaphorical meaning construction and learning outcomes.

2.3 Multimodal Annotation and Coding

Multimodal annotation was conducted using ELAN 6.7 (EUDICO Linguistic Annotator) to segment, label, and quantify multimodal resources in the selected microlesson videos. Following the stage segmentation criteria proposed by Zhang and Yuan, a six-stage instructional framework (lead-in, review, new study, practice, assignment, and summary) was adopted to align subsequent coding with pedagogical function [32].

A modality coding scheme was then applied, adapted from the multimodal classification proposed by Liu and Wang (see Table 2) [33]. Three modality categories were coded: auditory (A), visual (V), and kinesthetic (K). The auditory modality included teachers' speech (At), students' speech when present (As), and background music (Am). The visual modality included slide-based presentation (Vppt), material objects (Vo), and images (Vi). The kinesthetic modality captured embodied actions such as gestures (Kp), and body language (Kb). To enhance annotation reliability, three coders independently annotated the videos and resolved discrepancies through discussion to reach consensus before exporting the final statistics. Because the dataset consisted of single-presenter microlessons, learner participation occurred primarily through language-based instruction; accordingly, context-related variables such as Social distance (Es) and Individual distance (Ei) were not coded in the present study.

Table 2: Liu and Wang's Multimodality Classification

Modal type	Coding	Mode	Coding
Visual modality	V	Slide	Vppt
		Material object	Vo
		Images	Vi
Auditory modality	A	Teachers' speech	At
		Students' speech	As
		Background music	Am
Kinesthetic modality	K	Gesture language	Kp
		Body language	Kb
Environmental modality	E	Social distance	Es
		Individual distance	Ei

After completing the annotation process, ELAN statistical tools were used to extract the number of annotations and durations for each modality. These outputs were summarized using descriptive statistics and compared across microlesson series and instructional stages.

2.4 Learner Questionnaire and Expert Evaluation

A five-point Likert questionnaire (0 = strongly disagree, 4 = strongly agree), adapted from Shen's multimodal evaluation scale [34], was administered to 67 learners, all of whom had reached HSK Level 4 or above. Participants were randomly assigned to evaluate either Series I or Series II. 45 valid responses were retained. For each series, descriptive statistics were calculated across nine dimensions of perceived multimodal quality, including teachers' speech, on-screen text, images, gestures, facial expressions, video use, slide design, and background music. Independent-samples t-tests were conducted to examine between-series differences.

Seven experts from multiple universities evaluated the two microlesson series across auditory, visual, and kinesthetic dimensions. Expert feedback was summarized to identify strengths and weaknesses in multimodal design and to triangulate learner perceptions. The initial inter-coder agreement rate (percentage agreement) reached 90% across all annotated units. All discrepancies were subsequently resolved through joint discussion to reach a final consensus.

3. Results and Discussion

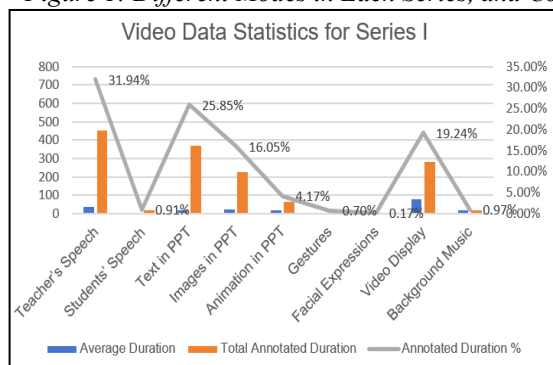
3.1 Multimodal Distribution and Interaction in TCFL Idiom Microlessons

This study analyzed 596 annotated segments from two series of award-winning idiom microlessons to examine the overall distribution and interaction of multimodal resources. As shown in Fig. 1 (a-d), both Series I and Series II share a similar distribution of multimodal resources. Teachers' speech, on-screen text, and images on slides constitute the three longest annotated durations, accounting for 31.94%, 25.85%, and 16.05% of the total annotated duration in Series I, and 29.96%, 37.40%, and 11.01% in Series II, respectively. By comparison, the use of students' speech, facial expression, and background music occurs only occasionally (Series I: 0.91%, 0.17%, 0.97%; Series II: 0.00%, 0.44%, 2.20%).

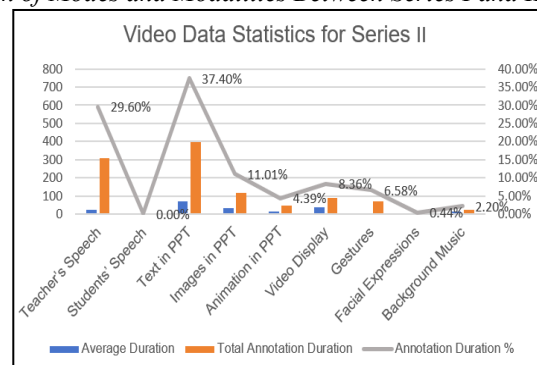
From the above data, it can be seen that both series of microlessons rely primarily on visual modalities, particularly text and images embedded in slide presentations. Through these visual cues, idiomatic meanings are anchored in concrete, familiar scenes, thereby facilitating learners' comprehension and cultural association. The auditory modality, represented by the teacher's speech, is also used frequently to explain, interpret, and elaborate on the definition or usage of idioms, thereby reinforcing conceptual understanding. Kinesthetic resources are relatively limited in both series. This is due to the structural nature of the microlesson format, which is largely monologic and screen-based, thus only leaving little space for interactive or embodied teaching behaviors. Nonetheless, despite their limited frequency, gestures are selectively employed at key points, especially when illustrating metaphorical mappings or spatially oriented idioms. These gestures serve as a valuable supplement to teachers' speech, enhancing teacher-learner engagement and strengthening students' understanding. The slightly higher kinesthetic proportion in Series II also suggests a more dynamic visual-embodied style of instruction, where the teacher actively integrates gesture to visualize metaphorical structures.

In addition to the above similarities, the two series of microlessons also demonstrate some different multimodal strategies. Series I rely more heavily on the auditory modality (37.83%) and video display (19.24%), whereas Series II shows lower proportions for auditory input (31.80%) and video display (4.39%). Meanwhile, the proportion of kinesthetic cues increases from 0.83% in Series I to 7.02% in Series II, suggesting a more dynamic visual–embodied instructional style in Series II. These kinesthetic cues, whether deictic or metaphorical, contribute to emphasizing key concepts, visualizing abstract idiomatic meanings, and maintaining discourse rhythm.

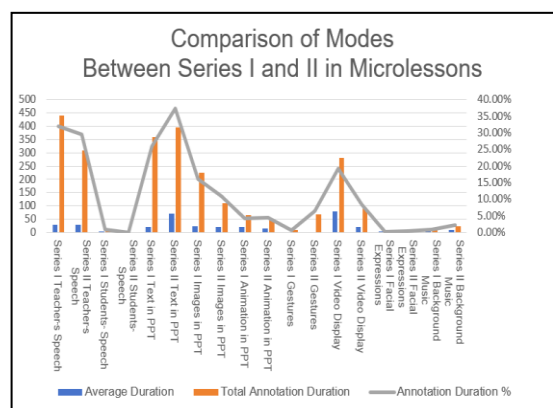
Figure 1: Different Modes in Each Series, and Comparison of Modes and Modalities Between Series I and II.



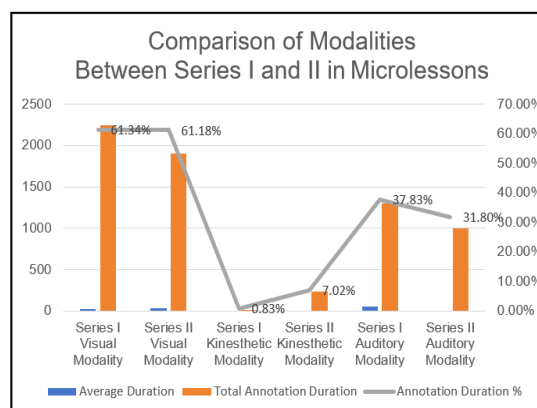
(a) Different modes in Series I.



(b) Different modes in Series II.



(c) Comparison of modes between Series I and II.



(d) Comparison of modalities between Series I and II.

According to Zhang's multimodal relations framework, the multimodal patterns observed in the two series of idiom microlessons primarily reflect complementary non-reinforced (C-nR) relations, in which visual and auditory modalities cooperate through coordination [7]. In both Series I and Series II, texts and images on slides illustrate idioms with definitions, examples, and contextual images, while the auditory modality, represented by teachers' speech (37.83% in Series I; 31.80% in Series II), explains these visual cues by clarifying figurative meanings, elaborating metaphorical associations, and contextualizing idiom use in discourse. Without either modality, the communicative intent of idiom teaching would be incomplete or difficult for international learners to comprehend.

This finding differs from the complementary-reinforced (C-R) relation typically observed in traditional face-to-face classrooms [35], where the teacher's speech serves as the primary modality and the visual modality plays a secondary role. In idiom microlessons, however, teaching is screen-based and pre-structured, with information visually foregrounded through on-screen text, images, and animations. As a result, meaning in TCFL idiom instruction is jointly constructed through the coordination of auditory and visual modalities, with kinesthetic modality reinforcing comprehension and learner engagement.

3.2 Modalities Across Instructional Stages in the Microlessons

According to Zhang and Yuan's multimodal discourse segmentation criteria, a six-stage instructional framework (lead-in, review, new study, practice, assignment, and summary) was adopted to examine how modalities are distributed and coordinated across pedagogical stages [32]. Based on the 596 annotated segments from both series, the further analysis examined how different modalities were distributed and coordinated across these stages.

As shown in Table 3, across most stages, the auditory and visual modalities exhibit C-nR relations, indicating that they cooperate to construct meaning without redundancy. The kinesthetic modality forms C-R relations with speech in all the teaching stages, serving to highlight key points, maintain teaching rhythm and enhance expressiveness. In the summary stage, the relation between auditory and visual modalities shifts to C-R, as teachers' speech becomes foregrounded and visual information functions mainly as reinforcement.

Table 3: Multimodal Relational Patterns in Different Teaching Stages in Series I and II.

Series	Stage	Auditory modality	Visual modality	Kinesthetic modality
I	Lead-in	Main modality	C-nR	C-R
	Review	Main modality	C-nR	C-R
	New study	Main modality	C-nR	C-R
	Practice	C-nR	Main modality	C-R
	Assignment	Main modality	C-nR	C-R
II	Lead-in	Main modality	C-nR	C-R
	New study	C-nR	Main modality	C-R
	Practice	C-nR	Main modality	C-R
	Assignment	C-nR	Main modality	C-R
	Summary	Main modality	C-R	C-R

Figure 2 presents the time proportion of modalities across stages, illustrating how multimodal coordination aligns with pedagogical focus. However, since the two series did not realize all six stages as discrete segments, only stages that were explicitly present in each video were coded. Specifically, Series I did not include a standalone summary stage, and Series II did not include a review stage; these stages were therefore not separately coded or reported in the stage-level analyses.

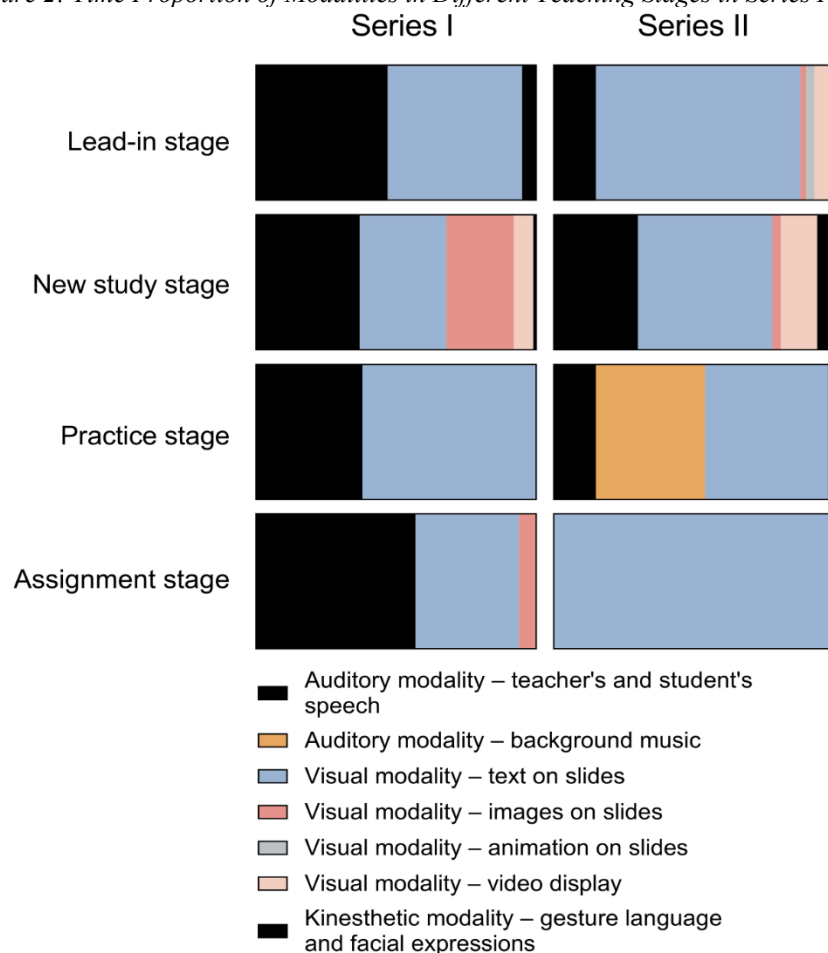
In the lead-in stage, teachers' speech remains the dominant mode, while on-screen text presents only brief keywords or thematic prompts. Gestures and facial expressions appear selectively to attract learners' attention and establish engagement. Both Series I and Series II demonstrate concise modal combinations, reflecting the principle of economy in initial input. Since idiom topics are introduced briefly to activate prior knowledge rather than convey detailed and new information, excessive multimodal layering is avoided to prevent cognitive overload.

In the new-study stage, which constitutes the core of idiom instruction, the two series differ in modal dominance. Although visual elements such as slide text and images remain visible throughout, in Series I they mainly serve as information carriers accompanying teachers' oral explanation. As shown in Table 3, the auditory modality thus functions as the main modality to meaning making. In Series II, visual modalities provide bilingual texts, illustrative images, and contextual examples to present idiom meanings and usage, while teachers' speech mainly connects and supplements these visual inputs.

In the practice stage, both series showed a visual dominant pattern. The practice stage is essential for students to internalize the learning content through hands-on practice. Therefore, it is crucial to guide students' enthusiasm and initiative and empower them to take ownership of their learning at this stage. Visual modality (text on slides) is used to explain the rules, while the auditory modality (teachers' speech) is used to guide the practice process. Slide-based animations and background music are used as supportive cues to sustain attention and facilitate participation, rather than to introduce new conceptual content.

In the assignment stage, Series I and Series II showed different patterns. Series I typically delivered homework through teacher's speech, with slides as support. Series II, by contrast, delivered homework through the slides, while teacher's speech served to emphasize the key requirements. In both series, assignment was delivered through a speech-slide combination, with each mode compensating for the other when needed.

Figure 2: Time Proportion of Modalities in Different Teaching Stages in Series I and II



In response to Research Question 1, the above analysis reveals that multimodal teaching in TCFL idiom microlessons is achieved through the coordinated integration of auditory and visual modalities, whose dominance varies with the pedagogical focus of each stage. The kinesthetic modality, though less frequent due to the screen-based format of microlessons, is strategically employed to support learners' comprehension of abstract or metaphorical meanings.

3.3 Multimodal Construction of Metaphorical Meaning in Idioms

Many Chinese idioms are metaphorically motivated, and idioms teaching inevitably involves mapping conceptual structures between source and target domains. From a cognitive perspective, idioms and metaphors share similar representational and inferential processes. The analysis of the two series of idiom microlessons shows that multimodal design plays a crucial role in realizing these metaphorical mappings, where visual, auditory, and kinesthetic modalities jointly externalize abstract cognitive associations and facilitate embodied understanding. Specifically, three major metaphor types are identified in the two series, animal metaphors, body metaphors, and mythological metaphors, each realized through distinctive multimodal strategies.

In animal metaphors, the visual modality provides concrete representations of the source domain, while the auditory modality elaborates their figurative meanings. Visual images, by depicting familiar animals and their behaviors, help learners activate prior knowledge and construct embodied associations between the source and target domains, thereby deepening their conceptual understanding of idiomatic meaning. For example, in teaching “井底之蛙” (“a frog at the bottom of a well”), images and animations in PPT illustrate the literal scene, and the teacher's explanation guides learners toward its extended metaphorical interpretation of limited perspective. This combination forms a C-nR relation, in which visual and auditory inputs cooperate to convey both surface and abstract meanings.

In body metaphors, teachers integrate gestures, facial expressions, textual and visual cues to guide learners in simulating embodied sensations. In this way, students can map bodily perceptions onto abstract meanings, thereby inferring the figurative implications embedded in physical actions. Kinesthetic cues form C-R relations with auditory explanations, enhancing emotional vividness, expressive intensity, and engagement in the comprehension process.

In mythological metaphors, images and short animations are employed to present historical and cultural contexts, supporting learners' understanding of idioms' origins and moral implications. Across all types, the integration of visual and auditory inputs facilitates learners' transition from perceptual familiarity with the source domain to conceptual understanding of the target domain.

From a pedagogical perspective, the multimodal construction of metaphorical meaning in TCFL idiom teaching can be summarized in three aspects. First, understanding metaphors requires transforming abstract concepts into concrete. This process is effectively achieved through multimodal coordination, where visual modalities such as images and on-screen text provide direct conceptual grounding, while auditory explanations elaborate on underlying figurative meanings, enabling learners to grasp metaphorical mappings between domains. Second, different modal strategies should be applied to the source and target domains: the source domain can be represented through clear visual cues, whereas the target domain requires precise verbal elaboration, particularly for idioms grounded in Chinese cultural contexts that demand deeper interpretation. Finally, each teaching stage serves distinct pedagogical purposes and therefore requires appropriate multimodal integration across stages. Overall, integrating multimodality with metaphor helps make abstract meanings more concrete, enhances cognitive engagement, and facilitates idiom comprehension in TCFL microlessons.

3.4 Evaluation of Multimodal Effectiveness in TCFL Idiom Microlessons

Student evaluation. A questionnaire survey was administered to 67 students, yielding 45 valid responses (response rate = 67.16%). Adapted from Shen's (2019) multimodal evaluation scale [34], it adopted a five-point Likert format (0=strongly disagree, 4=strongly agree) and showed satisfactory reliability (Cronbach's $\alpha > 0.80$, KMO = 0.68). Results indicated that 93% of respondents enjoyed Chinese classes and 64% preferred multimodal instruction. As shown in Table 4, Series I scored highest on video ($M = 3.09$) but lowest on facial expression ($M = 2.09$), whereas Series II scored highest on teachers' speech ($M = 3.27$) and lowest on images in slides ($M = 2.75$). Significant differences were found in facial expression ($t = -3.939$, $p < .001$) and body movement ($t = -4.053$, $p < .001$).

The results suggest that Series I, focusing on zodiac idioms, was rich in verbal and visual explanation but limited in embodied engagement—an approach that ensured clarity yet underused gestures that could have enhanced cultural and metaphorical understanding. In contrast, Series II relied heavily on speech and text, which, though coherent, appeared visually monotonous; contextual images might have made idioms more vivid and memorable. Overall, learners valued clear verbal delivery and structured visuals, but the findings highlight the need for a more balanced integration of auditory, visual, and kinesthetic modalities to optimize comprehension and engagement in TCFL idiom microlessons.

Table 4: Descriptive statistics of nine evaluated dimensions in Series I and II

Dimensions	Minimum	Maximum	Mean	SD
Series I				
teachers' speech	2	4	2.91	0.802
Text on slides	1	4	2.68	0.829
Body language	1	4	2.32	0.800
Facial expression	0	4	2.09	0.936
Images on slides	0	4	3.00	0.988
Font in slides	1	4	3.05	0.834
Teaching videos	1	4	3.09	0.884
Overall design of slides	1	4	2.89	0.813
Music and rhyme	0	4	2.64	0.917
Series II				
teachers' speech	1	4	3.27	0.758
Text on slides	2	4	3.2	0.701
Body language	1	4	3.00	0.778

Dimensions	Minimum	Maximum	Mean	SD
Facial expression	1	4	3.00	0.715
Images on slides	1	4	2.75	0.751
Font in slides	1	4	3.09	0.858
Teaching videos	1	4	3.16	0.680
Overall design of slides	1	4	3.05	0.806
Music and rhyme	1	4	3.00	0.863

Expert evaluation. To further examine multimodal design quality, seven experts from universities including Northwest Normal University, Lanzhou University, Dalian University of Foreign Languages, Guangxi Normal University, South China Normal University, Jilin University, and Zhejiang Normal University evaluated the two microlesson series based on visual, auditory, and kinesthetic dimensions.

Experts identified both strengths and weaknesses, emphasizing multimodal coordination as a key indicator of high-quality microlessons. Three main findings emerged from expert feedback.

(1) Speech-image coordination enhances comprehension. Experts emphasized that clear visuals, Chinese subtitles, and occasional English explanations in Series II facilitated learners' understanding and retention of idioms, suggesting that limited use of L1 support can aid comprehension in TCFL contexts.

(2) Auditory and kinesthetic modalities are central to evaluation. Experts agreed that the teachers demonstrated fluent and expressive speech delivery, complemented by natural gestures and facial expressions. However, minor issues such as overly fast pace and monotonous tone were noted, highlighting the need for rich intonation and moderate rhythm in microlesson narration.

(3) Moderation in video use is essential. Although embedded videos enriched microlessons, overlong segments were viewed as excessive. Quantitative analysis further indicates that video display accounted for about one-fifth of the total annotated duration in Series I (19.24%), whereas the proportion was much lower in Series II (4.39%). This pattern supports the experts' view that embedded clips can enrich microlessons, but they should remain concise and tightly aligned with the teaching objective. In summary, both student and expert evaluations confirm that effective multimodal design hinges on the complementary use of auditory, visual, and kinesthetic elements. High-quality TCFL microlessons should foreground the teacher's oral performance, integrate supportive visual cues and gestures, consider bilingual accessibility, and maintain a balanced, moderate use of multimedia resources.

These evaluation findings align with and extend prior work on multimodal instruction. Consistent with the principles proposed by Zhang [7], our results confirm that effective modality use adheres to effectiveness, adaptation, and economy. Unlike traditional face-to-face classrooms where teacher speech typically serves as the primary channel of instruction [7, 35], our findings demonstrate that in microlesson settings, visual modalities—particularly on-screen text, images, and video—assume a substantially greater and more active role in meaning construction. Royce [23] observed that multimodal resources in TESOL classrooms—particularly visual and auditory modalities—can be configured in various logical-semantic relations to construct coherent instructional meaning. Extending this line of inquiry to the microlesson context, our analysis reveals that complementary non-reinforced coordination (C-nR) constitutes the prevailing intermodal pattern in idiom microlessons—a distinction that reflects the screen-based, pre-structured nature of video-based instruction. Similarly, whereas Zhang and Wang [24] emphasized the importance of examining multimodal interaction in relation to communicative contexts, our study operationalizes this framework by specifying how modal configurations shift systematically across lead-in, new study, practice, and assignment stages in the microlesson format. Regarding kinesthetic modality, Wang [21] observed that gestures play a supporting role in Chinese immersion classrooms; our findings extend this observation by identifying their additional cognitive function in scaffolding metaphorical mapping for idiom comprehension. Together, these comparisons position the present study as both a validation and a contextual extension of existing multimodal frameworks, offering empirical, stage-sensitive evidence for designing effective video-based idiom instruction.

4. Pedagogical Implications for TCFL Idiom Microlesson Design

Drawing on the multimodal and metaphorical analyses above, three pedagogical implications can be proposed for the design of TCFL idiom microlessons.

(1) Emphasize multimodal complementarity rather than accumulation. In the microlesson format, multimodal design provides an efficient means of delivering linguistic and cultural knowledge within limited time, ensuring learner focus and interaction in online learning contexts. Teachers' speech and on-screen text function as the primary meaning carriers, while visual and kinesthetic cues provide complementary support through C–nR or C–R relations. Therefore, multimodality should not be pursued by merely increasing sensory channels, but by ensuring complementary coordination across modalities. For instance, clear visual cues can ground learners' understanding of the source domain, while teachers' verbal explanation extends these cues to metaphorical interpretation.

(2) Match modality use with teaching content and stage. Effective idiom instruction requires matching modality use with both the nature of the teaching content and the pedagogical focus of each stage. In the lead-in stage, concise auditory and visual input such as key phrases, images, and brief narration helps activate learners' prior knowledge and interest. During the new-study stage, teachers' speech should be closely coordinated with on-screen visuals, while gestures and facial expressions play a supportive role in explaining idiomatic meanings and guiding metaphorical interpretation, making metaphorical mappings easier to follow. In the practice stage, slide-based text typically carries task prompts, while teachers' speech guides the learning procedures. In the assignment stage, speech and slides usually combined to make task requirements explicit, with either mode taking the lead depending on task design.

At the design level, a comparison of the two microlesson series indicates that visual presentation should remain clear, coherent, and pedagogically purposeful. Text and images should complement rather than repeat each other, with images clarifying abstract meanings and concise text highlighting essential information. Similarly, auditory delivery, including tone, pacing, and intonation, should be expressive yet well controlled to ensure clarity and maintain attention. Matching modality use with instructional content and stage thus promotes cognitive efficiency, pedagogical coherence, and deeper understanding of idiomatic and metaphorical meaning.

(3) Integrate embodied cognition and contextual interaction across modalities. Beyond verbal-visual coordination, TCFL idiom microlessons should incorporate embodied and contextual interaction among modalities to deepen learners' metaphorical understanding. According to the theory of embodied cognition [36–38], human cognition is grounded in perceptual and bodily experiences, and abstract concepts are often understood through metaphorical extensions of sensorimotor schemas. Idioms such as “一马当先” (to take the lead) and “呆若木鸡” (dumbstruck with fear) exemplify this mechanism: the former maps bodily experience of movement and direction onto abstract notions of initiative, while the latter reflects a physical state of immobility as a metaphor for emotional paralysis.

In idiom microlessons, teachers can leverage gestures, facial expressions, and situational visuals to evoke such embodied simulations, enabling learners to feel metaphorical meanings through perceptual and motor resonance rather than verbal explanation alone. When these embodied cues are integrated with contextual auditory and visual inputs—such as narrative background, sound effects, or scene-based imagery—learners are guided to reconstruct idiomatic meaning within lived communicative scenarios. This multimodal–embodied approach not only supports semantic comprehension but also strengthens affective engagement and memory retention.

Such an approach is consistent with Zhang's multimodal collocation principles of effectiveness, adaptation, and economy, emphasizing that each modality should serve a distinct pedagogical function, be flexibly aligned with instructional goals, and contribute efficiently to meaning construction [7].

5. Conclusions

This study examined how multimodal resources are distributed, coordinated, and pedagogically deployed in TCFL idiom microlessons, with particular attention to stage-based multimodal relations and the multimodal construction of metaphorical meaning. By analyzing two award-winning microlesson series through ELAN annotation and multimodal relation coding and triangulating these patterns with learner questionnaires and expert evaluations, the study provides empirical evidence on how high-quality idiom microlessons orchestrate auditory, visual, and kinesthetic resources to support meaning-making.

Across both series, teachers' speech, slide-based text, and images constituted the core resources, and meaning was mainly built through complementary (non-reinforced) auditory–visual coordination. Kinesthetic cues were strategically employed at key explanatory moments to highlight metaphorical mapping, emphasize critical semantic contrasts, and maintain learner attention. Importantly, multimodal coordination varied systematically by instructional stage and content. Lead-in segments tended to use economical modality combinations to activate attention, whereas the new-study stage showed the greatest variability as idiomatic meanings were explained; practice stages generally used speech-supported visuals for recall and application; and the assignment stage more often used speech-slide combination to deliver the homework. These findings point to a complementary, functionally appropriate approach to multimodal design guided by adaptability, efficiency, and economy. In particular, in TCFL idiom microlessons, well-timed gestures can help learners move from literal cues to figurative meaning when idiomatic mappings need to be made explicit.

This study makes both theoretical and pedagogical contributions to TCFL multimodal instruction. Theoretically, it extends multimodal discourse analysis from traditional face-to-face classrooms to the microlesson context in three ways. First, by demonstrating that modal relations—particularly the distinction between complementary-reinforced (C-R) and complementary non-reinforced (C-nR) coordination—are not static but shift systematically across instructional stages, this study challenges the static view of modal configurations prevalent in previous research [7, 24]. Second, unlike prior studies that have largely focused on general modal distribution or static text-image relations, our stage-sensitive analytical framework offers a more dynamic and pedagogically-oriented model for examining video-based instruction. Third, by integrating metaphor theory with multimodal analysis, this study reveals how visual, auditory, and kinesthetic resources jointly externalize abstract cognitive associations and facilitate cross-domain mapping in idiom comprehension—a dimension largely underexplored in previous TCFL multimodal research. Pedagogically, the findings translate into three evidence-based design guidelines: (a) prioritizing complementarity over accumulation, ensuring each modality serves a distinct pedagogical function; (b) aligning modality choices with instructional stages and content types; and (c) integrating embodied cues such as gestures and facial expressions to scaffold metaphorical comprehension, particularly for culturally grounded idioms. These recommendations directly inform the design of more effective TCFL idiom microlessons in technology-enhanced language education, offering novice teachers and instructional designers a practical reference for multimodal resource orchestration.

Despite its contributions to understanding multimodal coordination and metaphor-based meaning construction in TCFL idiom microlessons, this study still has several limitations. First, the analysis was based on two prize-winning microlesson series, which may not capture the full range of instructional practices and production conditions in TCFL contexts. In addition, the questionnaire involved a relatively small sample size and relied on learners' perception instead of direct learning outcomes, which may limit the generalizability of the evaluation results. Future studies should expand both the scope of materials and the participant pool to build a more comprehensive picture of international learners. Such extensions would help test the robustness of the stage-specific multimodal patterns in this study and refine multimodal design guidance for TCFL idiom microlessons.

References

- [1] Wang, H. Y. (2018). Design of a Blended Smart Education Model for English Micro-lessons. *Teaching & Administration*, 18, 98–101.
- [2] Feng, F. (2022). Micro-lessons Promote Changes in Teaching Formats and Teachers' Roles. *People's Education*, 21, 122–123.
- [3] Shieh, D. (2009). These Lectures are Gone in 60 Seconds. *Chronicle of Higher Education*, 55(26), 1-13.
- [4] Hu, T. S. (2011). "Micro-lesson": A New Trend in the Development of Regional Educational Information Resources. *e-Education Research*, 10, 61–65. <https://doi.org/10.13811/j.cnki.eer.2011.10.020>
- [5] Han, J. L. (2019). Micro-Lecture Teaching for Improving the Learning Effect of Non-English Majors at North China Electric Power University. *English Language Teaching*, 12(6), 209-216. <https://doi.org/10.5539/elt.v12n6p209>

- [6] Conti, S. (2017). Chengyu in Chinese Language Teaching: A Preliminary Analysis of Italian Learners' Data. *Irish Journal of Asian Studies*, 3, 59-82.
- [7] Zhang, D. L. (2015). *Theory of Multimodal Discourse Analysis and Foreign Language Teaching*. Higher Education Press.
- [8] Jiang, P., Wijaya, T. T., Mailizar, M., Zulfah, Z., & Astuti, A. (2022). How Micro-lectures Improve Learning Satisfaction and Achievement: A Combination of ECM and Extension of TAM models. *Mathematics*, 10(19), 3430. <https://doi.org/10.3390/math10193430>
- [9] Xie, Q. (2022). The Factors Influencing Chinese University Teachers' Intentions for Using the Micro-lecture in the Post COVID-19 Era. *International Journal of Environmental Research and Public Health*, 19(22), 14887. <https://doi.org/10.3390/ijerph192214887>
- [10] Hadji-Daud, J. A., Gandamra, H. S., Salazar, D. A., Luib, J. R. T., Magsayo, J. R., & Malayao, S. C. (2025) Development of an Interactive Contextualized Microlesson on Correlation for Grade 11 Learners. *International Journal of Scientific Engineering and Science*, 9(5), 192-196
- [11] Liu, R., & Wang, H. (2014). Design and Practice of a Flipped Classroom Instructional Model Based on Micro-lessons. *Modern Educational Technology*, 24(5), 26–32. <https://doi.org/10.3969/j.issn.1009-8097.2014.05.004>
- [12] Wang, H. L. (2022). Research on Innovation of Informationized Models in College English Teaching in the New Media Era. *Information Science*, 40(8), 50–55.
- [13] Rui, H. L., Li, B. J., & Zhou, S. M. (2023). Evaluation of Teaching Methods and Pathways for Enhancing Effectiveness in Ideological and Political Education Courses in Higher Education. *China Higher Education*, 21, 37–40.
- [14] Wan, M. (2016). An Empirical Study of the Flipped Classroom Model in College English Teaching. *Higher Education Exploration*, 5, 69–72.
- [15] Li, J. (2023). Design Features of Near-synonym Discrimination Microlessons: Evidence from Prize-winning “Tang Feng Hanyu” Microlessons. *Chinese Character Culture*, 23, 20-23. <https://doi.org/10.14014/j.cnki.cn11-2597/g2.2023.23.009>
- [16] Wang, G., & Wang, Z. (2025). Characteristics of Culture-Themed Microlessons in International Chinese Education: Evidence from Award-winning Works of the 5th National Postgraduate Chinese Teaching Microlesson Competition. *Journal of Educational Institute of Jilin Province*, 41(6). 137-142.
- [17] Yu, Z. (2025). Teaching Chinese Grammar through International Chinese Language Education Micro-lectures: Negotiating Mass and Presence through Multimodal Pedagogic Discourse. *International Review of Applied Linguistics in Language Teaching*, 63(1), 121-145. <https://doi.org/10.1515/iral-2024-0050>
- [18] Nie, Y. (2023). Application of Multimodal Multimedia Information and Big Data Technology in Teaching Chinese as a Foreign Language Course. *International Journal of Digital Multimedia Broadcasting*, 1, 2257863. <https://doi.org/10.1155/2023/2257863>
- [19] O'Halloran, K. L. (2004). *Multimodal Discourse Analysis: Systemic-Functional Perspectives*. Bloomsbury Publishing.
- [20] Kress, G., & Van Leeuwen, T. (2020). *Reading Images: The Grammar of Visual Design*. Routledge.
- [21] Wang, Z. L. (2016). On Multimodal Discourse of Chinese Immersion Program in the United States. *Journal of Research on Education for Ethnic Minorities*, 27(4), 124–129. <https://doi.org/10.15946/j.cnki.1001-7178.2016.04.019>
- [22] Jewitt, C. (2012). *Technology, Literacy, Learning: A Multimodal Approach*. Routledge.
- [23] Royce, T. (2002). Multimodality in the TESOL Classroom: Exploring Visual-verbal synergy. *TESOL Quarterly*, 36(2), 191–205. <https://doi.org/10.2307/3588330>

- [24] Zhang, D. L., Wang, Z. (2016). Theoretical Framework for Multimodal Interaction Analysis. *Foreign Languages in China*, 13(2), 54–61. <https://doi.org/10.13564/j.cnki.issn.1672-9382.2016.02.008>
- [25] Wang, S., & Luo, H. (2021). Exploring the Meanings and Grammatical Functions of Idioms in Teaching Chinese as a Second Language. *International Journal of Applied Linguistics*, 31(2), 283-300. <https://doi.org/10.1111/ijal.12334>
- [26] Wu, C., & Wang, S. (2017). The Selection and Application of the Example Sentences of Idioms in International Chinese Language Teaching. *Current Research in Chinese Linguistics*, 96(1), 1–17.
- [27] Cucchiari, C., Hubers, F., & Strik, H. (2022). Learning L2 idioms in a CALL Environment: The Role of Practice Intensity, Modality, and Idiom Properties. *Computer Assisted Language Learning*, 35(4), 863-891. <https://doi.org/10.1080/09588221.2020.1752734>
- [28] Macedonia, M., Müller, K., & Friederici, A. D. (2011). The Impact of Iconic Gestures on Foreign Language Word Learning and Its Neural Substrate. *Human Brain Mapping*, 32(6), 982-998. <https://doi.org/10.1002/hbm.21084>
- [29] Macedonia, M. (2014). Bringing Back the Body into the Mind: Gestures Enhance Word Learning in Foreign Language. *Frontiers in Psychology*, 5, 1467. <https://doi.org/10.3389/fpsyg.2014.01467>
- [30] Zhang, H. (2016). *Idiom Representation and Processing: A Neurocognitive Approach*. Shanghai Foreign Language Education Press.
- [31] Gibbs, R. W., Jr., Bogdanovich, J. M., Sykes, J. R., & Dale, J. B. (1997). Metaphor in Idiom Comprehension. *Journal of Memory and Language*, 37(2), 141–154. <https://doi.org/10.1006/jmla.1996.2506>
- [32] Zhang, D. L., & Yuan, Y. Y. (2011). A Study of Modal Synergy in Dynamic Multimodal Discourse: A Case Study of Multimodal TV Weather Forecast Discourse. *Shandong Foreign Language Teaching*, 32(5), 8. <https://doi.org/10.3969/j.issn.1002-2643.2011.05.002>
- [33] Liu, J. Y., & Wang, S. (2019). Multimodal Discourse Analysis in the Pedagogy of an International Chinese Elementary Comprehensive Course. *International Journal of Chinese Language Education*, 5, 67–91.
- [34] Shen, W. (2019). *Application of multimodal teaching model to English reading instruction in higher vocational colleges (Master's thesis)*. Xinjiang University.
- [35] Zhang, D. L. (2009). On a Synthetic Theoretical Framework for Multimodal Discourse Analysis. *Foreign Languages in China*, 6(1), 24–30.
- [36] Gibbs, R. W., Jr. (2006). *Embodiment and Cognitive Science*. Cambridge University Press.
- [37] Lakoff, G., & Johnson, M. (1999). *Philosophy in the Flesh: The Embodied Mind and Its Challenge to Western Thought*. Basic Books.
- [38] Lakoff, G., & Turner, M. (1989). *More Than Cool Reason: A Field Guide to Poetic Metaphor*. University of Chicago Press.

Funding

This research was funded by Key Project of the 2026 Research Program of the Hangzhou Collaborative Innovation Research Center for Language Services (No. 26JD097)

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgment

This paper is an output of the science project.

Copyrights

Copyright for this article is retained by the author (s), with first publication rights granted to the journal. This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).