

## Original Paper

# Multimodal Persuasion Strategies in English Learning

## Advertisements

Yuan Gao<sup>1</sup>

<sup>1</sup> School of Foreign Studies, North China University of Water Resources and Electric Power, Zhengzhou, China

Received: April 12, 2026

Accepted: May 29, 2026

Online Published: June 11, 2026

doi:10.22158/eltls.v8n3p217

URL: <http://dx.doi.org/10.22158/eltls.v8n3p217>

### **Abstract**

*In the digital communication era, English learning advertisements have evolved from language-only persuasion to complex semiotic events that integrate language, image, sound, and space. Drawing on Kress and van Leeuwen's visual grammar and Aristotle's three persuasive appeals, this study analyzes two real-world cases—Duolingo's "social-first" campaign and Liulishuo's "AI Teacher" brand film—to examine how multimodal persuasion strategies operate in English learning advertisements. The findings reveal that these advertisements use language to create anxiety and promise transformation, visual images to construct ideal self-representations, sound to build emotional atmospheres, and interactive space to encourage user engagement. Through the orchestration of these modes, English learning advertisements achieve a distinctive persuasive mechanism: they redefine English proficiency not merely as a language skill but as a gateway to a higher-order identity. This identity reconfiguration drives consumer action. The study provides theoretical insights for advertising practitioners and contributes to the marketing research of English language education.*

### **Keywords**

*multimodal discourse analysis, persuasion strategies, English learning advertisements, visual grammar, identity construction*

## **1. Introduction**

On any social media platform, English learning advertisements are ubiquitous. With the rise of short videos, these advertisements are no longer simple text-and-image banners; they are complex ensembles of visuals, sounds, written language, and interactive elements designed to manipulate viewers' emotions and prompt purchases.

This communication phenomenon is grounded in multimodal discourse analysis (MDA). Since the

1990s, social semioticians such as Kress and van Leeuwen have argued that meaning is no longer confined to language but is co-constructed through the interplay of images, sounds, colors, and layouts (Kress & van Leeuwen, 2006; Kress, 2010). Such multisensory stimulation can more precisely mobilize audiences' emotions, establish credibility, and present facts, thereby enhancing persuasion.

English learning advertisements have a unique persuasive logic compared to other product categories. Unlike fast-moving consumer goods that appeal to sensory pleasure or car advertisements that signal social status, English learning advertisements sell a promise of capability enhancement and identity transformation: Learn English, and you will enter a wider world. This logic gives rise to highly patterned multimodal strategies—from generating anxiety about English ability to portraying an idealized future, from demonstrating scientific methods to triggering immediate action.

Using Kress and van Leeuwen's visual grammar and Aristotle's three appeals (ethos, pathos, logos), this study analyzes two representative campaigns—Duolingo's social-first marketing and Liulishuo's AI Teacher brand film—across four modal dimensions: language, visual, auditory, and spatial.

## 2. Theoretical Framework and Literature Review

### 2.1 Multimodal Discourse Analysis and Advertising

The theoretical foundation of MDA was laid by Kress and van Leeuwen in *Reading Images* (2006). Extending Halliday's (1994) three metafunctions of language, they proposed a visual grammar that analyzes images through representational meaning (narrative vs. conceptual), interactive meaning (contact, distance, perspective), and compositional meaning (information value, salience, framing). This framework remains the most widely used tool for multimodal advertising analysis.

In advertising research, multimodal studies have accumulated substantial findings. Feng and O'Halloran (2012) applied multimodal metaphor theory to reveal how visual metaphors underpin brand persuasion. Meng (2018) examined print English learning advertisements in China, demonstrating how multiple sign systems cooperate in multimodal discourse.

Internationally, research on English learning advertisements has grown. Oh (2022) combined critical discourse analysis and MDA to investigate how South Korea's top three online English learning platforms construct learners' incompetent identities, showing that these advertisements idealize native-speaker standards and reinforce Korean learners' self-deprecation. Karim and Haq (2024) analyzed IELTS preparation advertisements, revealing how language proficiency is commodified and how multimodal strategies shape consumer perception and experience.

However, existing multimodal advertising research has largely focused on fast-moving consumer goods, luxury brands, or public service announcements. Studies specifically targeting English learning advertisements remain relatively scarce, particularly regarding the dominant multimodal forms of short-video advertisements, social media advertisements, and interactive advertisements. Moreover, the literature tends to stay at the textual analysis level, with insufficient attention to the synergistic mechanisms among modes.

## 2.2 Research Questions

This study addresses three questions:

1. What persuasive functions do the linguistic, visual, auditory, and spatial modes respectively perform in English learning advertisements?
2. How do these modes collaborate synergistically?
3. What distinguishes the multimodal persuasion strategies of English learning advertisements from those of other product categories?

## 3. Case Studies

Two real-world cases are analyzed: (1) Duolingo's "social-first" campaign (2023-2025) and (2) Liulishuo's "AI Teacher" brand film (2018).

### 3.1 Duolingo's Social-First Campaign

#### 3.1.1 Campaign Background

Duolingo is one of the world's largest language learning platforms. Since 2021, it has adopted a social-first strategy, making social media content the smallest creative unit. The core methodology is "Doing whatever it takes to make you keep your streak." Duolingo's mascot, a green owl named Duo, plays the protagonist. The marketing team defines Duo's personality as "crazy, quirky, and weird," which has made the character deeply popular on Chinese social media as "sassy" and "dramatic."

#### 3.1.2 Linguistic Mode

Duolingo's language strategy differs radically from traditional English learning advertisements. Instead of a brand monologue, Duo speaks in the first person to users. In social media posts, Duo's language ranges from sarcastic scolding to pleading: "Are you really not studying?" "Please, just five minutes." This highly colloquial, emotional language blurs the boundary between brand and friend, constructing a personalized interactive relationship.

In the Obituary event, Duolingo's official account unexpectedly posted a black-and-white photo of Duo with a mourning caption. This linguistic move broke users' expectations, created strong cognitive dissonance, and drove massive shares and discussions. From Aristotle's perspective, this is a classic pathos appeal: humor and surprise trigger emotional responses that embed the brand in memory.

#### 3.1.3 Visual Mode

Duo's visual design is the core asset of Duolingo's persuasion. A round, green owl whose expressions seamlessly switch among pitiful, angry, proud, and depressed—this visual emoticon system is itself a persuasive language. On the in-app widget, Duo's pose and expression change multiple times a day, sometimes offering inspirational quotes, sometimes looking furious, all urging the user to maintain their streak.

From a visual grammar perspective, Duo's design has several persuasive features. First, its rounded proportions and simplified facial features conform to the baby schema preference, evoking viewer affection. Second, the brand's green color aligns with the metaphor of learning as growth. Third, Duo's

expression changes follow a predictable pattern: anger when the user doesn't study, happiness when they return. This visual reward and punishment mechanism externalizes the abstract act of learning into concrete emotional feedback—a form of visual persuasion.

#### 3.1.4 Multimodal Synergy

Duolingo's sophistication lies in the deep synergy between linguistic and visual modes. When Duo's angry face is captioned "Are you really not studying?", the text gives the expression a specific referent, while the expression turns the text into palpable emotion. The user is no longer a passive advertisement viewer but drawn into a relational game with Duo: Not studying brings guilt, studying brings the satisfaction of making Duo happy.

Moreover, Duolingo's push notifications, widgets, and social media content form a cross-platform multimodal persuasion network. Users see Duo's stalking expression on their phones, watch Duo memes on social media, and see streak badges shared by friends. This interweaving embeds language learning into daily life.

### 3.2 *Liulishuo's AI Teacher Brand Film*

#### 3.2.1 Campaign Background

Liulishuo is a pioneer in China's online English education. It launched the world's first AI English teacher in 2016, and by 2018 had accumulated over 80 million registered users. In 2018, it released a brand film titled "The Ambition of AI," featuring two AI teachers named Peter and Tony. The film included a trailer and three short stories: "High Intelligence," "The Low-Paid," and "AI Teacher's Daydream." It won the 2018 Longxi Creativity Award and offers exemplary multimodal persuasion design.

#### 3.2.2 Linguistic Mode

The film's core linguistic strategy is "Using complaints to highlight selling points." In the "High Intelligence" episode, Tony sees the slogan "Never rest, at your service" and complains, "Who do they think we are? Low-end robots?" Peter silently uses his laser eye to change the slogan to "High intelligence, teach you English!" In the "Low-Paid" episode, Tony complains, "Human teachers cost 500 an hour, but we are online 24 hours everyday for only 99 a month." This creates humor while precisely communicating the AI teacher's cost-effectiveness.

The genius of this complaint strategy is that the AI teacher's self-mockery (low pay, lack of respect) implicitly conveys three persuasive messages: 24 hours online everyday availability (convenience), far cheaper than human teachers (economy), and the AI teacher has emotions (humanization). From the three appeals, humor generates pathos, comparative numbers provide logos, and the AI's candid complaints build ethos.

#### 3.2.3 Visual Mode

The most noteworthy visual design is the AI teacher's appearance. According to Liulishuo's vice president, the creative team initially tried 3D modeling but found it too rigid; they switched to live actors with prosthetic makeup. The result preserved a machine-like stiffness while allowing human

facial micro-expressions and emotional changes. The colored light strip at the hairline changes color to symbolize the AI's emotional state.

This visual strategy achieves a semiotic balance. A purely mechanical AI would create distance; a purely human would fail to convey the AI concept. Liulishuo's solution finds the middle ground: non-human enough to signal technology, human enough to build emotional connection. In terms of visual grammar's interactive meaning, the AI teachers look directly at the camera and speak to the audience, creating a quasi-face-to-face teaching relationship—more persuasive than traditional voice-over narrations.

#### 3.2.4 Auditory and Audio-Visual Modes

The auditory design is equally polished. The AI teachers speak with mechanical pauses and slight electronic voice effects. This auditory marker signals their AI identity while creating a comedic contrast between the robotic voice and emotionally charged content. The mismatch between mechanical sound and emotional content is itself a source of humor, and laughter lowers resistance to advertising.

In terms of music and editing, each short story uses fast-paced dialogue with light background music. Each episode ends with a visual punchline (e.g., Peter splitting a desk in anger), releasing the comedic buildup. The brand slogan—We swear to improve humanity's English—is repeated at the end of each story.

#### 3.2.5 Multimodal Synergy

The Liulishuo film uses the linguistic mode to generate humor and memorability, the visual mode to balance machine and human, and the auditory mode to reinforce identity and comedic effect. Their synergy makes the abstract concept of an “AI English teacher” tangible. This resembles Duolingo's strategy in that both create a virtual character with personality and emotion to build user connection; but Liulishuo leans more toward rational persuasion (*logos*), while Duolingo leans more toward emotional drive (*pathos*).

### 4. Discussion

Synthesizing the case analyses and broader observations of English learning advertisements, we can summarize the persuasive functions of the four modes:

Linguistic mode primarily creates tension and triggers action. It offers solutions through transformation narratives and uses imperative sentences to prompt immediate behavior.

Visual mode constructs an ideal self-image. The before/after contrast, the confident global citizen, and cultural symbols of the English-speaking world collectively portray a “future you,” prompting the viewer to think, “I could become that”.

Auditory mode builds emotional atmosphere and imbues the brand with human-like warmth. In short-video advertisements, suspenseful sound effects during the problem-presentation phase and uplifting music during the solution phase create an emotional trajectory of tension and release. In Duolingo and Liulishuo, the voice quality of the character is key to personality construction.

Spatial mode transforms passive viewing into active participation. In H5 advertisements, in-app demos, and social media interactive content, actions like swiping, clicking, commenting, sharing, and streak-keeping become part of the persuasion process. Duolingo's streak-and-share mechanism turns learning into social display, leveraging social proof to reinforce user retention.

The persuasive mechanism of English learning advertisements goes beyond the problem-solution logic of ordinary product advertisements. It operates at the level of identity reconstruction. It does not simply say, "This product solves a specific problem"; it implies, "By learning English, you will become a more complete, more valuable person." As Oh (2022) sharply observed, such advertisements construct a wang-chobo (always deficient) learner identity—never good enough, always chasing native-speaker standards—thereby generating learning anxiety and fueling continuous consumption. The tension between this deficiency narrative and the empowerment narrative of the solution is the deep source of persuasive power in English learning advertisements.

## 5. Conclusion

Using multimodal discourse analysis as a theoretical lens and Duolingo's social-first campaign and Liulishuo's AI Teacher film as cases, this study systematically examined multimodal persuasion strategies in English learning advertisements. The main findings are:

1. Persuasion in English learning advertisements is a prototypical multimodal process. The four modes have distinct functions: language creates anxiety and narrative, visuals construct the ideal self, sound builds emotional atmosphere, and interaction guides user engagement.
2. The fundamental difference between English learning advertisements and other product categories is that they do not sell a commodity but a new self who speaks English. This identity-reconstruction discourse is made visible through the strategic deployment of multimodal resources.

This study has limitations. It focuses on only two brands, and the analysis is qualitative without empirical testing on the user side. Future research could expand the case range and use eye-tracking experiments, surveys, or interviews to verify the actual effects of multimodal persuasion strategies. Additionally, emerging technologies such as AI-generated advertisement content deserve further investigation into their impact on the multimodal persuasion landscape.

## References

- Feng, D., & O'Halloran, K. L. (2012). Representing emotive meaning in visual images: A social semiotic approach. *Journal of Pragmatics*, 44(14), 2067-2084.
- Forceville, C. (1996). *Pictorial metaphor in advertising*. Routledge.
- Halliday, M. A. K. (1994). *An introduction to functional grammar* (2nd ed.). Edward Arnold.
- Karim, A., & Haq, S. (2024). Marketing language proficiency: A multimodal discourse analysis of IELTS advertisements and their impact on consumers' perception and their experience. *Pakistan Journal of Social Sciences*, 44(1), 45-59.

- Kress, G. (2010). *Multimodality: A social semiotic approach to contemporary communication*. Routledge.
- Kress, G., & van Leeuwen, T. (2006). *Reading images: The grammar of visual design* (2nd ed.). Routledge.
- Meng, Y. (2018). *The construction of coherence in multimodal discourses: A study on the print English learning advertisements in China*. Shanghai Jiao Tong University Press.
- Oh, S. (2022). Marketing the Wang-chobo identity: Problematic construction of English learner identities in the advertising discourse of South Korea's English learning platforms. *The Journal of the Korea Contents Association*, 22(3), 123-136.
- Wiswall, S. (2024). *Duolingo's TikTok strategy: Entertainment-first approach to brand communication* (Undergraduate thesis). Elon University.