DOI: 10.53469/jssh.2025.8(02).12

The Effects of Subtitle Reading on Information Acquisition in Multimodal Text Viewing: An Eye-Tracking and Questionnaire-Based Study

Lin Chen[1,*], Ying Xia[2]

[1,2]College of Foreign Languages, Shanghai Maritime University, Shanghai 201306
[1]*chenlin@shmtu.edu.cn*, [2]*xiaying0087@stu.shmtu.edu.cn*
**Correspondence Author*

Abstract: ***With the rapid digitalization of media environments, multimodal audiovisual materials have become an increasingly important resource for information access and second language learning. This study examines the role of subtitles in information acquisition during audiovisual viewing by integrating eye-tracking measures with performance-based learning outcomes. A total of 108 Chinese university learners of English as a foreign language (EFL) participated in an eye-tracking experiment and completed post-viewing tests assessing information acquisition under subtitled and non-subtitled conditions. The results showed that learners in the subtitle condition achieved significantly higher overall learning outcomes than those in the no-subtitle condition. However, correlation analyses revealed no significant association between visual attention to subtitle regions and test performance. These findings indicate that while subtitles facilitate information acquisition at the group level, the amount of visual attention allocated to subtitles alone does not reliably predict individual learning gains. The study provides empirical evidence for the complex relationship between visual attention and learning outcomes in multimodal environments, contributing to cognitive research on audiovisual translation and offering pedagogical implications for the effective use of subtitles in multimodal language learning contexts.***

Keywords: Subtitle reading, Multimodal learning, Eye-tracking, Information acquisition, Cognitive reception.

1. Introduction

Advances in digital media technologies have fundamentally reshaped the contemporary media environment. Traditional media constraints have gradually been dismantled, and multimodal resources such as social media, short videos, live streaming platforms, and virtual reality devices have become major channels for information dissemination and acquisition. The development of internet and multimedia technologies has enabled language learners to access a wider range of online learning resources through multimodal media.

Audiovisual materials, which integrate auditory, visual, and textual modalities, provide contextualized and dynamic input that is particularly conducive to language learning. According to dual-channel processing accounts (e.g., Plass & Jones, 2005), the integration of linguistic input (spoken or written language) with visual representations (static or dynamic images) activates both verbal and non-verbal cognitive systems, enhances input processing, and promotes deeper cognitive engagement. With the growing popularity of short-form video content and the widespread use of audiovisual materials in L2 contexts, audiovisual translation--primarily in the form of subtitling and dubbing--has assumed increasing pedagogical and communicative importance. In particular, subtitles have become a central interface through which learners access and process audiovisual input.

Previous research has suggested that subtitle reading is associated with information uptake in audiovisual contexts (Bird & Williams, 2002; Diao et al., 2007). However, subtitle processing does not necessarily translate into effective learning or information acquisition. The relationship between visual attention to subtitles and learning outcomes remains theoretically complex and empirically under explored. It is therefore necessary to examine the underlying information-processing mechanisms and the factors influencing audience engagement with multimodal texts in order to better understand the cognitive dynamics involved.

Within the domains of Audiovisual Translation and Second Language Acquisition, the cognitive processes underpinning information acquisition during subtitle-assisted viewing have received relatively limited systematic investigation. From a cognitive perspective, multimodal input may induce split-attention effects (Moreno & Mayer, 1999) or increase cognitive load due to the simultaneous processing of multiple information streams. These possibilities underscore the need for empirical research examining learners' allocation of visual attention and its relationship to learning outcomes in subtitle-supported multimedia environments.

The present study adopts an interdisciplinary framework integrating Audiovisual Translation Studies, Cognitive Translation Studies, and Reception Studies. It shifts the analytical focus to the reception process, specifically the cognitive mechanisms underlying viewers' engagement with subtitled multimodal texts. By doing so, the study extends research on translation cognition beyond production-oriented perspectives and contributes to a more comprehensive understanding of audience processing in audiovisual contexts.

Methodologically, the study employs a triangulated research design combining eye-tracking experimentation and questionnaire-based assessment. Eye-tracking technology enables the real-time observation of viewers' visual attention

and processing patterns during dynamic text viewing, thereby providing fine-grained evidence of online cognitive activity. The questionnaire-based outcome measures assess participants' information acquisition, allowing for the examination of the relationship between visual attention and learning performance. Through this empirical approach, the study seeks to identify patterns of attentional allocation in subtitle-mediated multimodal environments and to clarify how such patterns relate to information uptake and language learning. The findings are expected to provide evidence-based insights for subtitle translation research and practice, ultimately contributing to the enhancement of subtitle quality and pedagogical effectiveness.

Specifically, the study addresses the following research questions:

a) What patterns of visual attention do viewers exhibit when engaging with subtitled multimodal texts?

b) How does subtitle translation influence viewers' information uptake during multimodal text viewing? Is there a relationship between visual attention and learning outcomes?

c) In what ways can subtitle translation be optimized to facilitate information acquisition in multimodal learning contexts?

2. Literature Review

Since the late 1980s, translation cognition research has increasingly adopted theoretical frameworks and empirical methods from psycholinguistics, cognitive science, and neuroscience. The introduction of process-oriented methodologies—such as think-aloud protocols, keystroke logging, eye-tracking, and neuroimaging—has enabled systematic investigation of translators' cognitive processes during source-target language conversion (Jakobsen & Alves, 2021). This methodological shift has led to substantial advances in understanding translation as a complex, multidimensional cognitive activity, with a strong emphasis on translation process research (Hurtado Albir & Alves, 2009; O'Brien, 2006, 2011; Muñoz Martín, 2016).

Despite these advances, empirical research in translation cognition has remained predominantly translator-centered. Far fewer studies have examined the reception side of translation, namely how translated products are cognitively processed and interpreted by readers or audiences (Kruger & Kruger, 2017). This imbalance is notable given that reader response has long been central to translation theory. Early functionalist and communicative approaches, such as Nida's (1964) dynamic equivalence and German functionalist theories, foregrounded the role of the target audience and communicative purpose. However, these reader-oriented theories were largely developed through conceptual argumentation and textual analysis, with limited empirical validation due to methodological constraints.

Recent developments in cognitive science and experimental techniques have made real-time investigation of readers' cognitive processing increasingly feasible. In response, Kruger and Kruger (2017) proposed the notion of cognitive reception of translation, emphasizing the need to empirically examine how translated texts affect recipients' cognitive processing. This perspective aligns with reception-oriented approaches in literary studies, which shifted scholarly attention from textual properties to reader engagement and meaning construction. Within translation studies, this shift implies moving beyond equivalence-based comparisons toward an examination of translation as a cognitively and culturally situated communicative event.

Within the broader field, reception research has largely focused on macro-level or theoretical readers, whereas empirical studies involving real readers at the individual level remain relatively limited. Methodological challenges—including difficulties in capturing real-time cognitive responses and establishing reliable measures of reception—have constrained empirical progress. Notably, audiovisual translation, particularly subtitling, has emerged as a domain in which such challenges can be effectively addressed. The multimodal nature of audiovisual texts, combined with the availability of eye-tracking technology, provides a viable context for observing readers' attentional allocation and processing behavior during translation reception.

Parallel to these developments, translation process research has begun to acknowledge the relevance of audience-oriented cognition. Studies drawing on cognitive psychology and neuroscience suggest that translators may anticipate readers' knowledge and needs during translation (e.g., Annoni et al., 2012), while eye-tracking research has demonstrated the potential of extending process-oriented tools to the investigation of reception (Hvelplund, 2017). Nevertheless, these studies primarily aim to illuminate translators' cognitive processes, leaving the cognitive mechanisms underlying translation reception under-theorized and under-investigated.

Taken together, existing research highlights both the importance and the current limitations of cognitive reception studies in translation. Empirical work remains scarce, particularly with respect to how audiences process translated multimodal texts and how such processing relates to information uptake and learning outcomes. Moreover, the integration of cognitive translation research with theories of multimedia learning and audiovisual processing remains insufficiently developed. Addressing these gaps, the present study adopts an interdisciplinary approach that combines insights from cognitive translation studies, reception research, and multimedia learning. By employing eye-tracking and outcome-based measures, the study seeks to empirically examine the cognitive reception of subtitles in audiovisual contexts, thereby contributing to a more comprehensive and empirically grounded understanding of translation cognition.

3. Methodology

3.1 Participants

A total of 108 Chinese university students learning English as a foreign language (EFL) participated in the experiment. Participants were recruited from three parallel classes enrolled in an English Interpreting course at a leading university in Shanghai, China. All participants were native speakers of Chinese and second-language learners of English. Although

they came from different academic majors, they were enrolled in the same course and shared comparable instructional backgrounds and learning objectives.

Participants were between 18 and 22 years of age. To reduce potential confounding effects associated with demographic variability (e.g., age, professional experience, and reading proficiency), participants with similar age ranges and educational backgrounds were selected. While the findings are primarily representative of this population, the potential generalizability of the results to other learner groups is addressed in the limitations section.

All participants reported normal or corrected-to-normal vision and confirmed that they had no prior exposure to the experimental video. To further control for possible effects of background knowledge, the post-viewing questionnaire included items assessing participants' familiarity with the video content. These measures ensured that observed differences in eye-movement behavior and learning outcomes could be attributed to the experimental manipulation rather than pre-existing knowledge.

Participants were informed that all data would be treated confidentially and used exclusively for academic research purposes. Written informed consent was obtained from all participants in accordance with the ethical approval granted for this study.

3.2 Instruments and Materials

Eye movements were recorded using an EyeLink 1000 Plus eye tracker (SR Research) with a sampling rate of 2000 Hz. Standard eye-movement measures, including fixation count, fixation duration, saccades, and gaze trajectories, were extracted to examine participants' allocation of visual attention during video viewing.

Learning outcomes were assessed using a post-viewing questionnaire designed to measure audiovisual comprehension and vocabulary acquisition. The questionnaire was administered immediately after the eye-tracking experiment to capture participants' learning performance following exposure to the experimental materials.

The experimental stimulus consisted of a 3-minute-and-46-second English audiovisual excerpt selected from a TED Talk entitled *An Everyday Danger*. The video presents a popular science topic related to allergies and features standard pronunciation, a moderate speech rate, and lexical difficulty appropriate for university-level EFL learners.

In accordance with the experimental design, Chinese subtitles were produced by the authors, who specialize in Chinese–English translation. Subtitle preparation was carried out using Aegisub, a widely used open-source subtitle editing software. Established subtitling conventions regarding segmentation, line length, characters per line, synchronization, and on-screen presentation were strictly followed (Díaz Cintas & Remael, 2007). Subtitles were centrally aligned at the bottom of the screen and generally limited to a maximum of two lines to ensure readability and visual clarity.

3.3 Procedure

Participants viewed the experimental video while their eye movements were recorded, followed by a questionnaire-based assessment of learning outcomes. To avoid inducing attentional bias, participants were not informed in advance that subtitle processing constituted the primary focus of the study.

Each participant was seated individually in front of a computer monitor and instructed to (1) rest their head on a chin-and-forehead support, maintaining an eye-to-screen distance of approximately 40-70 cm; (2) minimize large head movements during video playback to ensure accurate eye-tracking recordings; and (3) wear headphones to reduce external noise and distractions.

Prior to the experimental session, a standard nine-point calibration procedure was conducted for each participant to ensure tracking accuracy. Calibration was accepted when average and maximum deviations did not exceed 1.0 and 1.5 of visual angle, respectively (SR Research EyeLink, 2022). Upon successful calibration, the experimental video was presented and played continuously from start to finish, automatically terminating at completion. All participants viewed the video under identical viewing conditions and for the same duration.

Participants were randomly assigned to one of two conditions: a subtitle condition (Chinese subtitles present) or a no-subtitle condition (no subtitles). Following the eye-tracking session, participants proceeded individually to a separate laboratory room to complete the post-viewing questionnaire under the supervision of the research team.

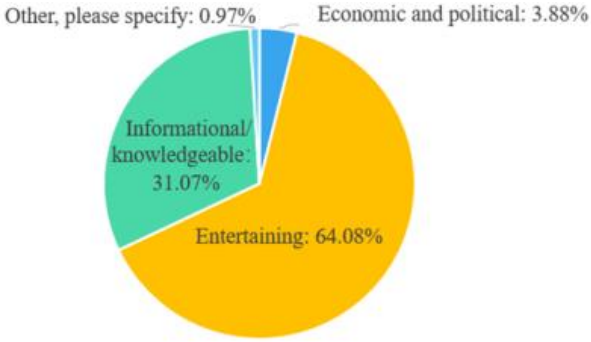
4. Results and Discussion

4.1 Questionnaire Results

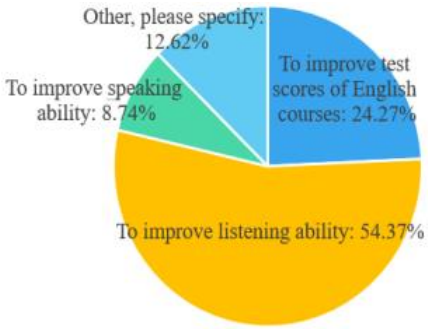
The first section of the questionnaire collected macro-level background information concerning participants' viewing habits and perceptions of audiovisual multimodal materials. Questions 1-5 examined participants' preferred types of audiovisual texts, motivations for viewing, weekly viewing duration, perceived comprehension barriers, and perceived advantages and disadvantages of using audiovisual materials for English learning. The aggregated responses are presented in Figure 1.

4.1.1 Viewing preferences and motivations

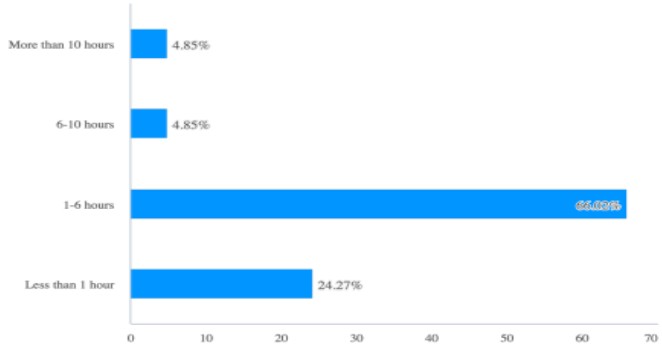
The results indicate that, among the 108 participants, entertainment-oriented audiovisual texts were the most frequently viewed category (64.08%), followed by informational or knowledge-oriented content (31.07%). Only a small proportion reported regularly viewing economic or political materials (3.88%) or other categories (0.97%). These results suggest that learners' exposure to English audiovisual input is largely shaped by entertainment-driven engagement rather than academic or professional purposes.



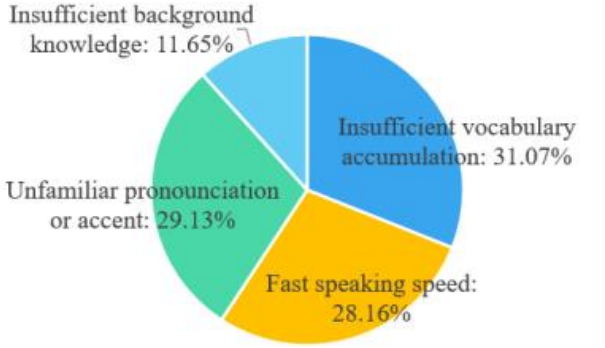
Question 1: Types of Audiovisual Texts Preferred by the Participants



Question 2: Motivation for Viewing Audiovisual Texts with Subtitles



Question 3: Average weekly duration of watching audiovisual materials for English learning



Question 4: Primary factors affecting the comprehension of audiovisual texts

Question 5: Advantages and disadvantages of using audiovisual materials for English learning

Figure 1: Feedback on Participants' Preferences for Audiovisual Texts

Regarding viewing motivation, more than half of the participants (54.37%) indicated that their primary goal was to improve listening comprehension, whereas a comparatively small proportion (8.74%) reported improving speaking ability as their main objective. Additionally, 24.27% viewed audiovisual materials to enhance academic performance in English-related courses and examinations. The prominence of listening-related goals underscores the perceived importance of auditory comprehension in English learning and supports the selection of audiovisual comprehension as a key indicator of information acquisition in the present study.

4.1.2 Viewing Duration

In terms of weekly engagement, 66.02% of participants reported spending between one and six hours per week using audiovisual materials for English learning. Only 4.85% reported six to ten hours, and an equivalent proportion reported more than ten hours per week. Although audiovisual materials are widely recognized as beneficial for language learning, the relatively moderate time investment observed here suggests that such resources function primarily as supplementary learning tools rather than primary instructional materials.

4.1.3 Perceived Comprehension Barriers

When asked to identify factors affecting comprehension of audiovisual texts, the largest proportion of participants (31.07%) identified limited vocabulary knowledge as the primary constraint. Comparable proportions reported unfamiliar pronunciation or accents (29.13%) and fast speech rate (28.16%) as major challenges. Only 11.65% regarded insufficient background knowledge as the main barrier.

These findings highlight the central role of lexical knowledge and phonological processing in audiovisual comprehension. The prominence of vocabulary limitations provides empirical support for including vocabulary acquisition as a second indicator of information uptake in this study, alongside audiovisual comprehension.

4.1.4 Perceived Advantages and Disadvantages

Participants generally reported positive attitudes toward audiovisual materials. Frequently mentioned advantages included increased engagement and interest, enhanced comprehensibility through multimodal integration, exposure to authentic contexts and natural speech rates, and support for

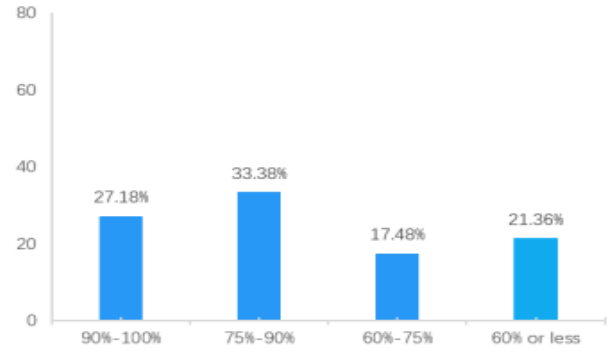
vocabulary development and listening practice. Several participants also noted that audiovisual materials reduce learning anxiety and promote memory retention through contextualized input.

At the same time, participants identified several limitations. A commonly reported concern was that learners may focus primarily on narrative content rather than linguistic form, potentially leading to superficial processing. Others suggested that audiovisual materials provide less explicit linguistic information than printed texts, which may limit opportunities for systematic learning. Some participants also emphasized the importance of careful material selection, noting that not all audiovisual resources are pedagogically appropriate.

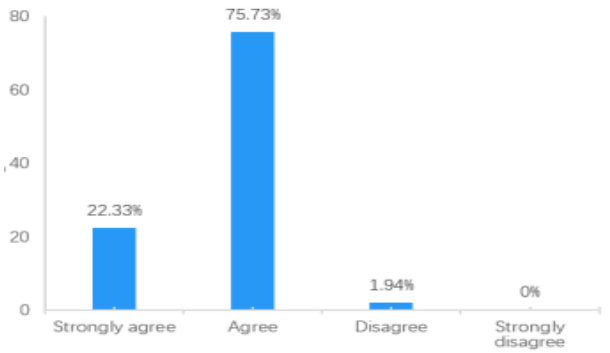
Overall, the questionnaire results indicate that participants frequently engage with audiovisual materials and hold generally positive perceptions of their pedagogical value. However, they also recognize lexical limitations and attentional challenges as potential constraints on effective learning. These self-reported perceptions provide important contextual background for interpreting the eye-tracking and learning outcome data presented in the following sections, particularly with regard to the relationship between visual attention allocation and information acquisition.

4.2 Self-Perceived Information Acquisition and Learning Outcomes

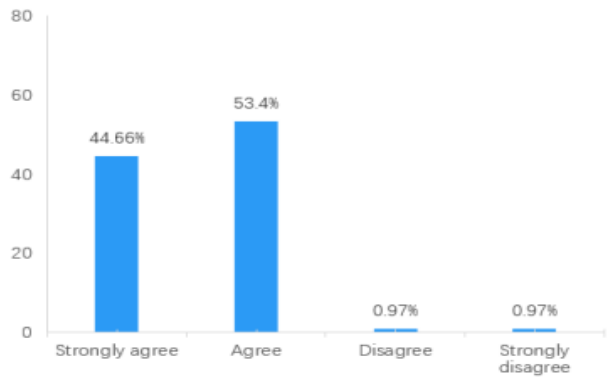
The second and third sections constituted the core components of the questionnaire. Section Two (Questions 6-10) examined participants' self-perceived information acquisition after viewing the experimental video materials. Responses were measured on a four-point Likert scale ranging from 1 ("strongly disagree") to 4 ("strongly agree"). The section comprised five items, yielding a maximum possible score of 20, with higher scores indicating stronger perceived information acquisition. The distribution of responses is presented in Figure 2.



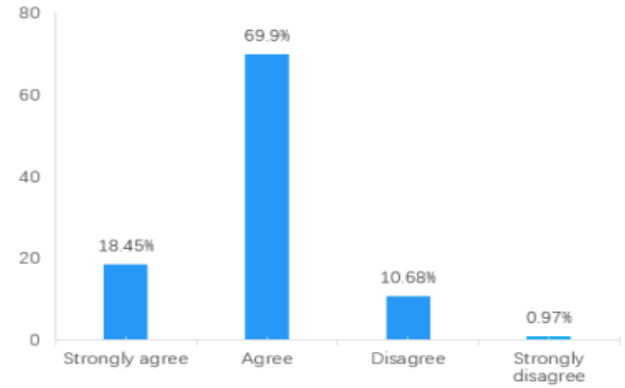
Question 6: Self-perceived comprehension of the experimental videos watched



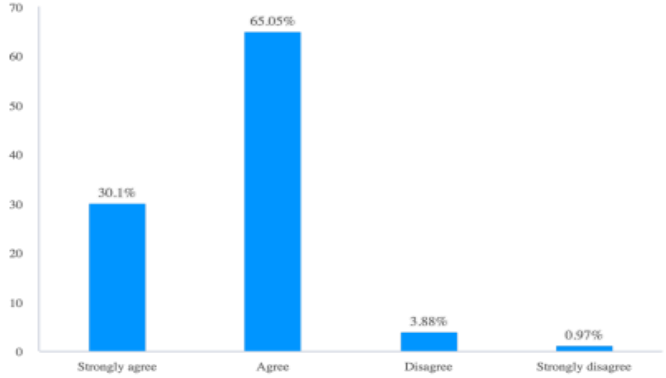
Question 7: Reading multimodal texts facilitates second language acquisition



Question 8: Reading multimodal texts facilitates audiovisual comprehension



Question 9: Reading multimodal texts facilitates vocabulary acquisition



Question 10: Reading subtitles facilitates audiovisual comprehension and vocabulary

Figure 2: Participants' Self-Perception after Viewing the Experimental Materials

4.2.1 Self-Perceived Audiovisual Comprehension

Regarding overall comprehension of the experimental videos, 27.18% of participants reported understanding more than 90% of the content, whereas 21.36% indicated that they comprehended 60% or less. The remaining participants reported intermediate levels of understanding (17.48% at 60–75%; 33.98% at 75–90%). Although comprehension levels varied, 78.04% of participants reported understanding more than 60% of the material, suggesting generally adequate processing of the audiovisual input at the group level. The relatively balanced distribution across comprehension bands further indicates substantial inter-individual variability, which is analytically relevant for subsequent comparisons between subjective perception and objective performance measures.

4.2.2 Perceived Pedagogical Value of Multimodal Viewing

Participants expressed overwhelmingly positive attitudes toward the role of multimodal text viewing in second language learning. A total of 98.06% either agreed (75.73%) or strongly agreed (22.33%) that multimodal materials facilitate second language acquisition. Similarly, 98% endorsed the view that multimodal viewing enhances audiovisual comprehension (44.66% strongly agree; 53.40% agree), indicating near-consensus regarding its comprehension benefits.

In contrast, endorsement was comparatively weaker for vocabulary acquisition. Although a substantial majority (88.35%) agreed or strongly agreed that multimodal viewing promotes vocabulary learning (18.45% strongly agree; 69.90% agree), a non-negligible proportion expressed disagreement (11.65% combined). This differential pattern suggests that learners perceive multimodal input as more directly supportive of global comprehension than of discrete lexical acquisition.

Participants also recognized the facilitating role of subtitles. A combined 95.15% agreed or strongly agreed that subtitle reading supports both audiovisual comprehension and vocabulary acquisition, while only 4.85% expressed disagreement. This strong endorsement aligns with theoretical accounts emphasizing the integration of visual-verbal channels in multimedia learning and provides an attitudinal backdrop for interpreting subtitle-related eye-movement patterns.

5.2.3 Learning Outcome Measures

The third section consisted of an objective learning outcome test designed to assess information acquisition. The test comprised 20 items: 10 measuring audiovisual comprehension and 10 assessing vocabulary acquisition. Each item was worth four points, thereby totaling 80 points. The relationships between learning outcomes and eye movements are examined in Section 4.3.

4.3 Information Acquisition and Eye Movements

Prior to statistical analyses, the eye-tracking data were screened for missing values and extreme outliers. Participants with invalid eye-movement recordings (e.g., zero fixation duration) were excluded. Extreme values exceeding ±3 standard deviations were removed. The final dataset consisted of 103 participants, including a subtitle group (PC) and a no-subtitle group (P).

Analysis were conducted on learning outcome measures and eye-movement indices, including fixation count, total fixation duration, interest-area (IA) dwell time, and interest-area fixation count. Descriptive statistics were first calculated for learning outcomes in both groups. Means, standard deviations, and ranges are reported in Table 1.

4.3.1 Descriptive Statistics and Group Differences in Leaning Outcomes

As shown in Table 1, the subtitle group (PC) obtained a higher mean learning score than the no-subtitle group (P). To examine whether this difference was statistically significant, an independent-samples t-test was conducted.

Levene's test indicated that the assumption of homogeneity of variance was met ($F = 2.228$, $p = .139$). The t-test results revealed a significant difference in learning outcomes between the two groups, $t(101) = -4.303$, $p < .001$. The mean difference was −10.99, with a 95% confidence interval ranging from −16.06 to −5.92, indicating that learners in the subtitle condition achieved significantly higher learning scores than those in the no-subtitle condition (Table 2).

Table 1: Descriptive statistics

	N Statistic	Minimum Statistic	Maximum Statistic	Mean		Std. Deviation Statistic	Variance Statistic
				Statistic	Std. Error		
Q-score P	56	18	89	65.95	1.913	14.318	204.997
Q-score PC	47	51	98	76.94	1.604	10.997	120.931

Table 2: Independent samples test of learning outcomes

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Q-score P	Equal variances assumed	2.228	.139	-4.303	101	<.001	-10.990	2.554	-16.057	-5.923
Q-score PC	Equal variances not assumed			-4.402	100.263	<.001	-10.990	2.497	-15.943	-6.036

4.3.2 Tests of Normality

Prior to examining the relationship between eye-movement measures and learning outcomes, tests of normality were conducted to determine the appropriateness of subsequent statistical procedures. As shown in Table 3, both IA dwell

time and IA fixation count in the subtitle group (PC) significantly deviated from a normal distribution according to the Shapiro–Wilk test (IA dwell time: W = 0.878, p < .001; IA fixation count: W = 0.924, p = .005), indicating substantial skewness in subtitle-related eye-movement measures.

In contrast, for the no-subtitle group (P), neither fixation count (W = 0.967, p = .126) nor fixation duration (W = 0.983, p = .620) showed significant departures from normality. Given the non-normal distribution of key eye-movement variables in the subtitle group, non-parametric statistical analyses were adopted for subsequent correlational analyses.

Table 3: Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
IA dwell time PC	0.165	47	0.003	0.878	47	0.000
IA fixation count PC	0.114	47	0.165	0.924	47	0.005
Fixation count P	0.094	56	.200*	0.967	56	0.126
Fixation duration P	0.071	56	.200*	0.983	56	0.620

a. Lilliefors Significance Correction

4.3.3 Correlations between Eye-Movement Measures and Learning Outcomes

Spearman rank-order correlation analyses were conducted to examine the relationships between eye-movement measures and learning outcomes within each group. The results are presented in Table 4.

Table 4: Correlations between eye-movements metrics and learning outcomes

Eye movement metrics		N	Q-score PC	Q-score P
IA dwell time PC	Spearman Correlation	47	-0.194	
	Sig. (2-tailed)		0.191	
IA fixation count PC	Spearman Correlation	47	-1.141	
	Sig. (2-tailed)		0.345	
Fixation count P	Spearman Correlation	56		-0.139
	Sig. (2-tailed)			0.307
Fixation duration P	Spearman Correlation	56		0.019
	Sig. (2-tailed)			0.887

*. Correlation is significant at the 0.05 level (2-tailed).

Within the subtitle group (PC), neither IA dwell time nor IA fixation count was significantly correlated with learning outcomes (r = −.19, p = .191; r = −.14, p = .345, respectively). Although both correlations showed a negative tendency, the relationships did not reach statistical significance. Similarly, within the no-subtitle group (P), no significant associations were found between learning outcomes and global eye-movement measures. Neither fixation count (r = −.14, p = .307) nor fixation duration (r = .02, p = .887) was significantly related to learning performance.

Taken together, the results indicate that individual differences in eye-movement behavior were not linearly associated with learning outcomes in either condition. The absence of significant correlations suggests a dissociation between online visual processing, as indexed by eye-movement measures, and offline learning outcomes. While subtitles significantly influenced learners' visual processing patterns and were associated with higher overall learning scores at the group level, the amount of visual attention allocated to subtitles alone did not reliably predict individual learning performance.

These findings underscore the non-linear and multifaceted nature of language learning from audiovisual input, in which visual attention, cognitive processing, and learning outcomes do not necessarily exhibit a direct one-to-one correspondence.

5. Conclusion

This study investigated the role of subtitle reading in multimodal audiovisual learning by integrating eye-tracking measures with objective learning outcomes. The results demonstrated that the presence of subtitles significantly enhanced overall information acquisition, as learners in the subtitle condition achieved higher post-viewing comprehension and vocabulary scores than those in the no-subtitle condition. These findings provide robust empirical evidence for the facilitating effect of subtitles in audiovisual learning among Chinese EFL learners.

At the same time, within the subtitle condition, no significant associations were found between subtitle-related eye-movement indices (interest-area dwell time and fixation count) and learning performance. This pattern reveals a dissociation between online visual attention and offline learning outcomes. While subtitles improved learning at the group level, the quantity of gaze allocated to subtitle regions did not reliably predict individual gains. The findings therefore suggest that subtitle-mediated learning involves complex cognitive integration processes in which visual attention represents only one component of a broader processing system. Attention allocation, processing efficiency, and multimodal integration likely interact in non-linear ways, rather than exhibiting a direct one-to-one correspondence with performance outcomes.

Theoretically, this study advances cognitive research on audiovisual translation by shifting the analytical focus from production-oriented processes to reception-based processing and learning consequences. It contributes to ongoing discussions regarding the relationship between eye-movement indicators and educational outcomes, highlighting the need to interpret gaze measures as indices of processing dynamics rather than direct proxies for learning success. Methodologically, the triangulation of eye-tracking metrics and performance-based assessments demonstrates the importance of combining online and offline measures to capture the multifaceted nature of multimodal learning.

Pedagogically, the findings support the use of subtitles as an effective scaffold in multimedia language learning environments. However, they also caution against evaluating subtitle effectiveness solely on the basis of visual attention measures. Increased fixation on subtitles does not necessarily imply deeper learning, underscoring the need for principled integration of subtitles within instructional design.

Several limitations warrant consideration. The participant sample consisted of university students with relatively homogeneous backgrounds, which may constrain generalizability. The study also employed a single audiovisual text of limited duration. Future research could examine learners of varying proficiency levels, different subtitle types (e.g., intralingual vs. interlingual), diverse task demands, and more fine-grained processing measures to further elucidate the mechanisms underlying subtitle-assisted learning.

Despite these limitations, the present study provides empirical evidence for the facilitative yet cognitively complex role of subtitles in multimodal language learning and contributes to a more nuanced understanding of the relationship between visual attention and learning outcomes in audiovisual contexts.

Acknowledgement

This study is funded by China Foreign Language Education Foundation (Project No. ZGWYJYJJ12A100).

References

[1] Annoni, J.-M., Lee-Jahnke, H., & Sturm, A. (2012). Neurocognitive Aspects of Translation. *Meta*, 57(1), 96-107.

[2] Bird, S. A. & Williams, J. N. (2002). The effect of bimodal input on implicit and explicit memory: An investigation into the benefits of within-language bubtitling. *Applied Psycholinguistics*, 23(4), 509-33.

[3] Diao, Y., Chandler, P., Sweller, J. (2007). The effect of written text on comprehension of spoken English as a foreign language. *The American Journal of Psychology*, 120(2), 237-261.

[4] Díaz Cintas, J. & Remael, A. (2007). *Audiovisual Translation: Subtitling*. Manchester, UK: St. Jerome Publishing.

[5] Hurtado Albir, A. & Alves, F. (2009). Translation as a cognitive activity. In J. Munday (Ed.), *The Routledge Companion to Translation Studies*. Oxon, New York, USA and Canada: Routledge.

[6] Hvelplund, K.T. (2017). Eye Tracking in Translation Process Research. *In The Handbook of Translation and Cognition* (eds J.W. Schwieter and A. Ferreira).

[7] Jakobsen, A. L. & Alves, F. (2021). Introduction. In F. Alves & A. L. Jakobsen (Eds.), *The Routledge Handbook of Translation and Cognition* (pp. 1-22). Oxon and New York: Routledge.

[8] Kruger, H. & Kruger, J. L. (2017). Cognition and reception. In J. W. Schwieter & A. Ferreira (Eds.), *The Handbook of Translation and Cognition* (pp. 71-89). Hoboken: John Wiley & Sons, Inc.

[9] Moreno R. & R. Mayer. (1999). Cognitive principles of multimedia learning: The role of modality and contiguity. *Journal of Educational Psychology*, 91(2), 358-368.

[10] Muñoz Martín, R. (2016). *Re-embedding Translation Process Research*. Amsterdam: John Benjamins.

[11] Nida, E. (1964). *Toward a Science of Translating with Special Reference to Principles and Procedures Involved in Bible Translating*. Leiden: E. J. Brill.

[12] O'Brien S. (2006). Eye-tracking and translation memory matches. *Perspectives-studies in Translation Theory and Practice*, 14(3), 185-205.

[13] O'Brien, S. (Ed.). (2011). *Cognitive Explorations of Translation*. London: Continuum.

[14] Plass, J. & Jones, L. (2005). Multimedia learning in second language acquisition. In R. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (pp. 467-488). New York: Cambridge University Press.