

The eye-ear relationship: investigating auditory impacts on subtitle reading and comprehension

A relação olho-ouvido: investigando os impactos auditivos na leitura e compreensão de legendas

Hussein Abu-Rayyash ^{*1}, Shatha Alhawamdeh ^{†2} and Yuri Ringomon ^{‡1}

¹Kent State University, College of Arts and Sciences, Department of Modern & Classical Language Studies, Kent, Ohio, USA.

²College College of Arts and Sciences Department of Middle East and North Africa Studies Oberlin, Ohio, USA.

Abstract

The study employs eye-tracking measures and comprehension quizzes to investigate linguistic processing during subtitle reading and comprehension in Arabic and Spanish. Specifically, it examines the impact of native English (L1) audio, foreign Spanish or Arabic (L2) audio, or the absence of audio on subtitle reading strategies and the processing of embedded linguistic units in novice learners. The study utilizes a 2x3 within-subjects design, manipulating audio condition (L1 audio, L2 audio, no audio) and subtitle language (Spanish, Arabic). Global eye-tracking analyses reveal subtitle reading patterns, while local analyses focus on the processing of specific linguistic units. Comprehension quizzes assess learning outcomes after the viewing of the subtitled videos. The findings suggest that the absence of audio leads to increased fixation durations, indicating greater attentional demand. Spanish subtitles result in higher fixation counts and prolonged reading times compared to Arabic, reflecting deeper cognitive processing. The study highlights the complex interplay between audio presence, subtitle language, and individual proficiency in shaping attentional allocation and lexical processing during language acquisition through audio-visual media. **Keywords:** Eye Movement. Audio. Subtitles. Reading. Comprehension.

Abstract

O estudo emprega medidas de rastreamento ocular e questionários de compreensão para investigar o processamento linguístico durante a leitura e compreensão de legendas em árabe e espanhol. Especificamente, examina o impacto do áudio nativo em inglês (L1), áudio estrangeiro em espanhol ou árabe (L2) ou a ausência de áudio nas estratégias de leitura de legendas e no processamento de unidades linguísticas inseridas em aprendizes iniciantes. O estudo utiliza um delineamento intra-sujeitos 2x3, manipulando a condição de áudio (áudio L1, áudio L2, sem áudio) e a língua das legendas (espanhol, árabe). As análises globais de rastreamento ocular revelam padrões de leitura de legendas, enquanto as análises locais se concentram no processamento de unidades linguísticas específicas. Os questionários de compreensão avaliam os resultados de aprendizagem após assistir aos vídeos legendados. Os resultados sugerem que a ausência de áudio leva a um aumento na duração das fixações, indicando uma maior demanda atencional. As legendas em espanhol resultam em um maior número de fixações e tempos de leitura prolongados em comparação com o árabe, refletindo um processamento cognitivo mais profundo. O estudo destaca a complexa interação entre a presença de áudio, a língua das legendas e a proficiência individual na alocação de atenção e no processamento lexical durante a aquisição de línguas através de mídia audiovisual.

Keywords: Movimento Ocular. Áudio. Legendas. Leitura. Compreensão.

Texto livre
Linguagem e Tecnologia

DOI: 10.1590/1983-3652.2024.52687

Session:
Articles

Corresponding author:
Hussein Abu-Rayyash

Section Editor:
Daniervelin Pereira
Layout editor:
João Mesquita

Received on:
May 21, 2024
Accepted on:
June 22, 2024
Published on:
November 4, 2024

This work is licensed under a
“CC BY 4.0” license.



1 Introduction

Audiovisual translation (AVT) has emerged as an essential tool for enhancing the accessibility of multimedia content to diverse audiences across languages (Díaz-Cintas; Hayes, 2023). Subtitling, a key component of AVT, involves displaying spoken dialogues and sounds on screen using synchronized

*Email: haburayy@kent.edu

†Email: salhawam@kent.edu

‡Email: yringomo@kent.edu

written text snippets (Fernández-Costales; Talavan; Tinedo, n.d.). While conventional subtitling presents the translation in the viewer's native language, reversed subtitling offers a less common approach by displaying foreign language subtitles alongside the original audio in the viewer's native language (Bolaños García-Escribano; Ogea Pozo, 2023). For language learners, especially those in the early stages of acquisition, watching subtitled movies can support vocabulary development by coordinating reading and listening. However, managing multiple information streams simultaneously may tax limited cognitive resources, potentially hindering comprehension, particularly for novice learners (Hornero Corisco; Gonzalez-Vera; Beltrán, 2023).

This study introduces a novel approach to investigating subtitle processing by employing a web-based eye-tracking system, allowing for remote data collection and expanding the potential participant pool beyond traditional laboratory settings. The present study employs eye-tracking measures and comprehension quizzes to investigate linguistic processing during subtitle reading and comprehension in two distinct languages: Arabic and Spanish. Unlike previous research, this study uniquely examines the impact of L1 (English) audio, L2 (Spanish/Arabic) audio, and no audio on subtitle reading and comprehension in novice learners of Spanish and Arabic, providing insights into the role of audio in subtitle processing across different language pairs. Specifically, vocabulary recognition assessments are used to measure learning outcomes after the viewing of the subtitled videos.

The study's within-subjects design allows for a more rigorous comparison of the different audio conditions, controlling for individual differences and enhancing the reliability of the findings. The researchers aim to shed light on the factors influencing subtitle reading strategies and the processing of various linguistic units in novice bilingual learners by comparing attentional coordination for vocabulary encoding across languages and cognitive manipulations. To achieve a comprehensive understanding of subtitle processing, the study conducts both global and local analyses of eye movements. This dual-level analysis approach, combining global subtitle reading patterns with local linguistic unit processing, represents a methodological advancement in the field of AVT research.

The global analysis examines reading patterns across entire subtitles, considering measures such as average fixation duration, total number of fixations, saccade length, and percentage of skipped subtitles. These measures offer insights into general reading approaches and the extent of reliance on subtitles in the absence of L1 audio support. Complementing the global analysis, the local analysis focuses on specific linguistic units embedded in the subtitles, such as verbs, nouns, adjectives, adverbs, expressions, phrases, sentences, and questions. Assessing first fixation duration, gaze duration, total reading time, and the probability of refixations or regressions on these linguistic units allows for the investigation of the processing and integration efficiency of various language components.

A key innovation of this study is the integration of eye-tracking measures with comprehension outcomes, providing a more empirically-based understanding of the relationship between attention allocation and learning in subtitle-supported language acquisition. The presence or absence of L1 or L2 audio is expected to differentially impact both global subtitle reading fluency and the processing of these linguistic units, which are crucial for language acquisition. Fixation and regression count on target linguistic units serve as indicators of overall integration success (Holmqvist *et al.*, 2011).

A primary objective of the study is to determine whether the presence or absence of L1 or L2 audio leads to greater dependence on subtitles for attentional allocation and the processing of linguistic units. There are still many who believe that audio-visual outlets for learning such as film or music don't set a good stage for learning, criticizing audio-visual outlets as being more distracting than traditional lecture style teaching. For instance, BORRÁS and LAFAYETTE (1994, p. 61) indicated that:

opponents of the use of subtitled video in foreign/second language teaching argue that the presence of subtitles is distracting and that they slow down the development of learners' listening abilities. Proponents of subtitles, on the other hand, contend that subtitles may help develop language proficiency by enabling learners to be conscious of language that they might -not otherwise understand.

This research hopes to install more high beams of evidence to lengthen the bridge between language learning and audio-visual stimulation. The research also explores potential differences in the impact

of audio presence on the processing of various linguistic units for more transparent Spanish compared to more opaque Arabic. Fixation and regression counts provide validation of overall integration success. In addition to the primary objective, the study investigates the links between subtitle reading fluency metrics, as quantified by eye tracking, and vocabulary learning outcomes, measured through recognition tests administered after the viewing of subtitled videos. Combining these measures allows the study to capture the multiliteracy impacts of audio manipulations on language acquisition. The research apparatus involves using RealEye, a webcam eye-tracking software, during authentic subtitled video viewing, paired with L2 vocabulary recognition assessments, aligning outcomes with specific research questions, and elucidating the contributions of language proximity and audio support to literacy development in novice learners.

The analysis of quantitative gaze metrics and comprehension scores reveals how language proximity and cognitive load influence attentional coordination efficiency, the processing of linguistic units, and overall understanding. These findings underscore the importance of auditory factors and language proficiency in subtitle design and delivery. To facilitate engagement and comprehension, content creators and educators should use native-language audio while supporting viewers with varying language expertise. The insights gained from this research have significant implications for language instruction and educational material design used throughout the language education field. In addition, the personal post survey is valuable to scholars of translation, of which, the cognition focused researchers are experiencing a time of flourishing of research interest and advancement. The results suggest tailoring instructional strategies to the specific needs and characteristics of the language being learned. For Arabic learners, using L2 subtitles without audio is highly effective, capitalizing on their ability to process visual information effectively. In contrast, Spanish learners may benefit from a balanced approach incorporating both L2 audio and subtitles, leveraging the linguistic alignment between the two modalities. These insights highlight the importance of considering each language's unique features and demands when developing instructional plans and resources.

The findings of this study have the potential to enrich theoretical perspectives on the coordination of input streams during complex literacy tasks and to inform the development of pedagogical guidelines that leverage multimedia reversed subtitling to support novice language growth, in line with multimedia learning principles. This research contributes to the growing body of knowledge in the AVT field and language acquisition, as well as translation studies by shedding light on the factors influencing subtitle reading strategies, the processing of linguistic units, and comprehension in novice bilingual learners.

2 Literature Review

Enhancing students' reading comprehension skills has emerged as a significant concern in both academic and non-academic settings (Kim, 2023). Translation plays a crucial role in enhancing reading comprehension, as noted by Alaboud (2022), who found that "translation could be an effective instructional strategy in improving learners' skills in reading comprehension in an EFL setting" Alaboud (2022, p. 424). In a case study of medical students, Rushwan (2017) indicated that the use of translation can facilitate and enhance reading comprehension skills of ESP medical learning. Concerning the utilization of subtitles to improve reading skills, several studies (Omar; Mohammed Salman; Vadivel, 2023; Qazi; Khan, 2023) have endorsed the notion that audio-visual materials enriched with subtitles manifest to enhance both second language (L2) reading comprehension abilities. For instance, Haider and Al-Salman (2022) indicate that the effects of viewing captioned videos on EFL learners revealed a positive impact on reading comprehension. Xu, Jordan, and Li (2022) also state that utilizing videos serves as a potent and engaging educational resource for EFL learners. In their study, Xu et al. provide further insights into the impact of viewing captioned videos on the listening and reading skills of university EFL learners. While the reading of subtitles has garnered increasing attention in recent decades, our comprehension of the cognitive processes involved in subtitle reading, and the variances from reading static text, remains constrained (Baranowska, 2020). Research in subtitle reading has been significantly influenced by the intersection of auditory input and visual processing. The relationship between what viewers hear and what they read on screen presents intriguing questions about cognitive workload and language comprehension.

To take in new information, our eyes must adjust their focal direction to where it is most concerned next, “in order to process information most effectively we must move our eyes so that the fovea fixates the location of that which we intend to process” (Schotter; Rayner, 2012, p. 84). One cannot presume that eye engagement behavior remains static throughout the language learning endeavor (Blythe *et al.*, 2006). Schotter and Rayner (2012, p. 86) emphasize that “Interestingly, the amount of disparity [between individual eye focus] tends to be greater in beginning readers than skilled readers”. Thus, our research seeks to tackle this trend early in the language learning process for second language learners. The pace at which they can become more fluent may differ and be faster than the pace they had when they learned their first language; however, such learners would be susceptible to individual eye focus disparity. Subtitles presentation through the purposeful utilization of colors, size, spacing, font and timing could have a varying degree of impact when considering these factors. In addition, video and subtitle speed is also an impactful factor upon the viewer. According to Liao, Yu, and Kruger (2020), as subtitle speed increases, word frequency and word length effects become less pronounced in local eye movement measures marking lexical processing. The authors conclude that increasing subtitle speed results in a shift “from local (cognitive) eye-movement control towards heuristics informed by global task constraints (e.g., subtitle speed)” Liao, Yu, and Kruger (2020, p. 430). This suggests that attention allocation during subtitle reading relies less on characteristics of individual words when subtitles are presented more rapidly.

Studies by Ross and Kowler (2013) and Szarkowska and Gerber-Morón (2018) highlight a trend where the presence of audio, especially in a viewer’s native language, leads to reduced reliance on subtitles. This phenomenon suggests that audio can support and enhance the subtitle reading experience, particularly when it is semantically aligned with the text. Another key consideration to be aware of is that the overall “reading process could be affected by semantically relevant auditory input in the context of reading English/L2 subtitles in video” (Liao; Yu; Kruger, 2020, p. 259).

Liao, Yu, Kruger, and Reichle (2022) present evidence that in situations involving multimodal reading, like reading subtitles in videos, eye movements are not solely regulated by visual information. Rather, readers simultaneously incorporate inputs from both visual and auditory modalities in the moment to determine when, where, and even whether to shift their eyes for subtitle reading. Thus, Liao, Yu, Kruger, and Reichle (2022) contribute to this discourse by examining how auditory input, even when partially redundant, influences the comprehension of subtitled content. Their study moves beyond the mere tracking of eye movements to consider how the audio-visual synergy affects viewers’ higher-level understanding of content.

A gap still exists in understanding the specific dynamics of how auditory context facilitates subtitle reading. V. (2020) partially addressed this by investigating second-language subtitles with native language audio. Yet, the absence of a no-audio condition in Ragni’s research limits the ability to discern the exclusive effects of auditory input on subtitle processing.

Given the widespread use of subtitles in second language education, understanding how audio influences subtitle comprehension is critically important. A range of prior research, as summarized by Liao, Yu, Kruger, and Reichle (2022), has studied the impacts of native and foreign language audio on global subtitle reading approaches using eye-tracking measures with languages like Dutch, Swedish, and English. Other work also revealed the effects of varied audio backing on patterns indicative of reading strategies (Szarkowska; Gerber-Morón, 2018). Additionally, recent studies measured local lexical and perceptual processing through fixation metrics (Bisson *et al.*, 2014; V., 2020).

Condinho Bravo (2008) extensively examined audiovisual-based language learning. In her work, Bravo emphasizes that subtitles should not be viewed as a panacea for foreign language acquisition. Despite the wealth of research on how audiovisual configurations affect second language learning, this topic remains relevant to related disciplines, particularly linguistics and translation studies, which in turn influence second language acquisition. Historically, translation has been employed in second language instruction by formal educational institutions. Furthermore, individual language learners have used translation as a method to test and improve their comprehension of the target language.

The current study builds upon these precedents by isolating the impacts of native English audio, foreign Spanish/Arabic audio, and no audio conditions during Spanish and Arabic subtitle reading

specifically for novice learners. The investigation goes beyond prior global analyses to align macro reading strategies with granular dynamic time courses of vocabulary integration efficiency. Through coordinated examination of sentence-level and localized word-level processing, the work comprehensively evaluates the contributions of audio presence and language proximity factors to literacy development gains.

While BORRÁS and LAFAYETTE (1994) found that the opportunity of having subtitles has a positive impact on comprehension, they also claim that it improves the productive use of the foreign language. A factor that is not underscored within this particular study, though it is of great consequence upon second language acquisition, is individual writing skill. In writing, phrasal usage is promoted as being a strategy that enhances the quality of a text. This is not only true for English, but also stressed in the Arabic writing system (Anis; Malik Abdul and S etiyadi, 2022). The videos used in this experiment certainly had phrases, and the participants may or may not have caught on to the phrases. Anis and Malik Abdul and S etiyadi (2022) found that the translations of Arabic literature (primarily poetry, in their example) can be done more naturally if the Arabic source text has phrases. Perhaps this finding then would also support a notion that suggests that phrasal usage in oral speech in videos could potentially be noticed by the listeners even if the language is their L2 language.

This eye-tracking study explores the impacts of native English language (L1) versus foreign Spanish or Arabic language (L2) auditory input, or its absence, on Spanish or Arabic subtitle reading and comprehension. The research utilizes a 2x3 within-subjects design with native English-speaking learners, manipulating audio condition (L1 English audio present, L2 Spanish or Arabic audio present, or no audio) and evaluating the effects on Spanish or Arabic subtitle processing.

The study aims to answer three questions:

1. How do L1 English audio, L2 Spanish or Arabic audio, and no audio conditions influence global subtitle reading efficiency patterns and local lexical processing in novice learners?
2. What is the relationship between L1 English or L2 Spanish or Arabic auditory input and comprehension accuracy with Spanish or Arabic subtitled foreign language content?
3. Does the presence or absence of L1 English versus L2 Spanish or Arabic audio impact perceived cognitive load during Spanish or Arabic subtitle reading?

3 Methodology

This section presents an overview of the methodology used in the study. It includes details about the sample and data collection and analysis processes. The study employs quantitative measures to investigate subtitle reading and comprehension in novice learners of Spanish and Arabic. Quantitative data is gathered through eye-tracking measures and comprehension quizzes to analyze participants' reading patterns and learning outcomes. Our methodological approach differs, though not to a radical degree, from that of Martínez and Gomez (2020), who found that competency in oral understanding of L2 intake increased through the usage of audio-visual tools and streaming services that use them. The following subsections of section 3 will provide more detail on the methodology.

3.1 Participants

The study involves 30 participants, comprising 20 native English-speaking university students as beginner learners of Spanish and 10 beginner learners of Arabic. All participants have studied their respective language (Spanish or Arabic) for only 1-2 semesters at the university level. To ensure a broad and unbiased participant pool, the recruitment phase involves outreach to language instructors. Participants are aged between 18-40 years to control for age-related cognitive and oculomotor changes. In accordance with Institutional Review Board guidelines, informed consent is obtained, along with assurances of data confidentiality and ethical handling of participant information. A pilot test with 5 participants is conducted to refine the study methodology, ensuring clarity in instructions and overall procedure.

3.2 Design

The study employs a 2x3 within-subjects design to assess the impact of subtitle language combined with auditory input. The study manipulated audio condition (L1 audio, L2 audio, no audio) and subtitle language (Spanish, Arabic). This design features 2 levels of subtitle language (Spanish and Arabic) crossed with 3 types of audio exposure (L1 English Audio, L2 Matched Audio, No Audio). Each native English participant, either a Spanish learner or an Arabic learner, completes 3 total viewing conditions specific to their language of study (L2: Spanish/Arabic) (see Table 1). To ensure the reliability of the results and counter potential biases, the sequence in which participants encounter these conditions is randomized, controlling for order effects and fatigue. This strategy is crucial for reducing learning or adaptation effects across sessions.

Eye-tracking data were collected using the RealEye webcam-based system, and comprehension was assessed through quizzes. In each condition, eye-tracking metrics are carefully analyzed to understand how participants interact with subtitles under varying auditory conditions. The study focuses on comparing these metrics within each language group across the audio conditions, aiming to quantify the linguistic challenges introduced by different audio inputs.

Data analysis involved both global and local analyses of eye movements. Global analysis examined reading patterns across entire subtitles, while local analysis focused on specific linguistic units. Key variables of interest include global subtitle reading strategies and local lexical processing efficiency, which are evaluated both within each language group and across different audio conditions. Statistical analyses, including ANOVAs and t-tests, were conducted to assess the effects of audio condition and subtitle language on eye-tracking metrics and comprehension scores.

Table 1. Study Conditions in a 2x3 (Language x Audio Condition) Within-Subjects Design

Condition	Subtitle Language (L2)	Audio Language
1	Spanish	Spanish
2	Spanish	English
3	Spanish	No Audio
4	Arabic	Arabic
5	Arabic	English
6	Arabic	No Audio

Source: Own elaboration.

3.3 Material

The study utilizes three one to two-minute video clips from the 2022 animation film *Pinocchio* available on Netflix, with the clips counterbalanced across audio conditions using a Latin Square design. Using clips from the same film controls for variability in factors like emotional engagement or audio-visual (AV) complexity that could influence processing beyond the linguistic components (Winke; Gass; Sydorenko, 2013). Counterbalancing ensures that differences found in eye-tracking metrics can be attributed to the audio manipulation rather than the intrinsic features of the videos. The content features simple conversational dialogues accessible to novice learners, between two main characters. The simplified narratives target novices while offering vocabulary learning opportunities. Drawing balanced samples from a common film maximizes internal validity within the controlled presentation system tailored to evaluate auditory impacts on second-language subtitle reading competency development.

The clip's duration provides an adequate volume of subtitles for analysis. The subtitles across the Spanish and Arabic video clips are comparable in length (average of 42.6 characters including spaces and punctuation), number of lines (one line used for all subtitles), and readability level based on Coh-Metrix metrics matching simplified conversational dialogues (Graesser *et al.*, 2014). This duration sustains participant concentration without inducing fatigue over the experiment. Netflix content ensures consistency in production quality and vocabulary level using the platform's filters. Subtitle parameters of font, size, and timing are standardized to eliminate visual variability.

3.4 Apparatus

Participants' eye movements are recorded monocularly using the RealEye webcam-based system. This system, as detailed by Lewandowska (2019), leverages advanced AI algorithms for accurate eye tracking, operating at a sampling rate of about 30 Hz, scalable up to 60 Hz under optimal conditions. The stimuli are displayed on a laptop screen, effectively simulating real-world online viewing contexts. The reason why the software was used exclusively on a laptop was because laptops provide more flexibility for eye-camera balance. Another reason was that the software itself operated better on a laptop than on a monitor, in which the eye-tracking software seemed to malfunction more. The RealEye system's distinctive feature of functioning entirely within a web browser, without requiring software downloads, enhances its accessibility and ease of use in remote settings. Subtitles are displayed in mono-spaced Courier New font, using 18pt white text with a 1pt black outline for visibility, ensuring readability across various viewing conditions. The data filtering process includes only fixations exceeding a 100ms duration within a 35px dispersion threshold, which is optimized for online processing and aligns with the RealEye system's capabilities (Lewandowska, 2019).

3.5 Procedure

The study utilizes RealEye's webcam tracking functionality for remote eye tracking, streamlining the process for participants by eliminating the need for software downloads. Instead, a shareable weblink grants access to the customized portal on personal laptops and desktops, supporting mainstream Windows and MacOS devices. The use of personal devices ensures comfort and familiarity for participants, although mobile devices are not suitable due to their size limitations. Upon entry into the study portal, participants receive clear, user-friendly instructions to guide the setup process, including assistance in positioning both the device and themselves for optimal capture within the webcam's view. Visual feedback confirms suitable face capture, especially the eyes, which may require minor adjustments in seating or lighting. Once the setup is satisfactory, participants initiate a calibration procedure, designed to be brief yet thorough, ensuring accurate tracking.

The procedure involves a randomized sequence presenting 35 distinct samples, each displaying a tracking dot against a neutral background. Participants are instructed to maintain focus and precisely click the target, with the webcam recording gaze spot estimates at each click. A 200-pixel deviation filter is applied between logged coordinates to ensure data accuracy, discarding any data points indicative of distraction or improper technique. To further validate the quality of data and participant engagement, a comprehension quiz is administered. The quizzes were designed on the target linguistic units and were conducted with the participants immediately after finishing the eye-tracking experiment. The utilization of vocabulary tests after the eye-tracking of subtitles is not novel, as Bisson *et al.* (2014) also implemented this method in their study, concluding with interesting findings that further cement subtitle integration's positive efficacy. They found that vocabulary retention was higher with subtitles than without, and it is anticipated that this study's results may corroborate such findings. However, this study links the comprehension score to the global and local eye-tracking measures.

3.6 Analysis

The analytical approach encompasses both generalized perspectives through global subtitle processing examinations and targeted dynamic evaluations of linguistic unit integration efficiency. Concurrently, these multi-level analyses aim to explain the complex interplay of factors influencing subtitle usage and literacy development under varied cognitive constraints. Specifically, the RealEye software pre-processes the raw gaze data recorded during full-length viewing trials, leveraging sophisticated algorithms to accurately classify fixations, saccades, and blinks. Velocity thresholds combined with spatial dispersion criteria are optimized for the parameters of the online experiment context. Subsequently, the resulting scanpath data files are exported to Python for quantitative aggregation and evaluation using the OpenGazeAnalyzer toolkit.

For global analyses, areas of interest (AOIs) are designated over the spatial bounds of subtitle text segments in all video frames. The global AOIs effectively encompass entire texts of subtitles, excluding

non-text screen regions. Various temporal indicators are then calculated over these aggregate areas, providing insights into high-level reading strategies. Metrics such as time to first fixation quantify attentional prioritization and capture of subtitle content across conditions. Meanwhile, total fixation duration and counts reflect overall reliance on subtitles for comprehension, with increased values indicating greater dependence in the absence of supportive auditory context. Furthermore, regression rates signify re-reading frequency, with higher rates implying difficulty in integration necessitating more repetitions. In essence, these metrics offer macro-level perspectives into attentional allocation patterns during complex literacy tasks under varied language proximity and cognitive load constraints.

Table 2. Target Linguistic Units for the three Experiments.

Linguistic Unit	English	Spanish	Arabic
First Experiment – L1 English audio and L2 Spanish/Arabic Subtitles			
Verb	to fear	Le da miedo	يخاف
Noun	war	guerra	الحرب
Adjective	weak	débil	ضعيف
Adverb	sometimes	a veces	أحياناً
Expression	they love you	te quieren mucho	يحبونك
Phrase	every day	todos los días	كل يوم
Sentence	You will see that I am not a coward.	Le demostraré que no soy cobarde.	سأثبت له أنني لست جبناً.
Question	Are you afraid?	¿Tienes miedo?	هل أنت خائف؟
Second Experiment – L2 English audio and L2 Spanish/Arabic Subtitles			
Verb	to lose	perdió	فقد
Noun	child	niño	طفل
Adjective	painful	doloroso	مؤلم
Adverb	very much	mucho	كثيراً
Expression	imperfect parents	padres imperfectos	آباء غير كاملين
Phrase	for a change	para variar	على سبيل التغيير
Sentence	It is something painful that you have to bear	Es algo doloroso que vas soportando	إنه شيء مؤلم يجب أن تتحمله
Question	What is a burden?	¿Qué es una carga?	ما هو العبء؟
Third Experiment – No audio and L2 Spanish/Arabic Subtitles			
Verb	go	Iré	أذهب
Noun	book	libro	كتاب
Adjective	proud	orgulloso	نغفور
Adverb	sometimes	a veces	أحياناً
Expression	I love it	¡Me encanta!	أحبه
Phrase	very good	muy bueno	مميز جداً
Sentence	I am going to be like Carlo	Voy a ser igual que Carlo	سأكون مثل كارلو
Question	Are you ready for school?	¿Listo para la escuela?	هل أنت مستعد للمدرسة؟

Source: Own elaboration.

An important part of the experiment set parameters was to make sure that the target words that the participants tested on would vary in parts of speech. This would then increase the difficulty of the task itself so as not to make stimulation too lulled. There were also phrases included. Table 2

above provides a list of the previously mentioned target words.

Concurrently, local dynamic analyses involve designating precise AOIs and isolating the set of pre-selected linguistic units embedded within the subtitles. As enumerated in Table 2, these linguistic units constitute key lexical items from various grammatical categories that crucially avoid overlap between the Spanish and Arabic conditions. Over these local AOIs, precise temporal indicators capture the subtle dynamics of lexical access and integration efficiency over time. Measures such as first fixation duration and gaze duration quantify the initial perceptual fluency of lexical activation when readers first encounter the critical words. Meanwhile, total reading time incorporates later processing, indicating the full duration of lexical access efficiency. Additionally, revisit ratios and regression probability reflect integration difficulty, with higher values marking regressions back to words posing encoding challenges. Essentially, these local metrics reveal the detailed time course over which vocabulary representations are constructed under the varied constraints imposed by auditory backing presence and language proximity factors. The local lexical processing efficiency analyses thus provide targeted dynamic evaluations, complementing the global generalized perspective.

Finally, supplementary content comprehension questions administered after each viewing session serve to confirm that attentiveness was sustained throughout the experiment. Accuracy scores help validate the eye-tracking data and conclusions drawn while aligning with specific research questions investigating multimedia learning phenomena. The combination of global and local eye-tracking analyses, along with comprehension assessments, provides a comprehensive framework for understanding the intricate dynamics of subtitle processing and language acquisition in novice bilingual learners. This multifaceted approach enables the exploration of the impact of auditory input and language proximity on attentional allocation, lexical processing efficiency, and overall comprehension, ultimately contributing to the development of evidence-based strategies for enhancing language learning through audiovisual media.

4 Results

4.1 Global Eye-Movement Analyses

This section presents the results of the global eye-movement analyses, examining reading patterns across entire subtitles. The analysis considers measures such as average fixation duration, total number of fixations, saccade length, and percentage of skipped subtitles. These measures provide insights into general reading approaches and the extent of reliance on subtitles under different audio conditions.

4.1.1 Average Fixation Duration

The analysis uncovered a main effect of the audio condition on participants' average fixation duration, $F(2, 18) = 12.34$, $p < .001$, $\eta^2 = .58$. Pairwise contrasts (see Table 3) showed that when no audio was present, fixation durations were significantly longer ($M = 507.75$ ms, $SD = 147.02$) compared to conditions with either L1 audio ($M = 290.25$ ms, $SD = 111.94$) or L2 audio ($M = 300.22$ ms, $SD = 111.11$), $ps < .01$. Furthermore, no significant differences were found between fixation durations for L1 and L2 audio conditions, $p > .05$. Moreover, the subtitle language did not significantly affect fixation duration, nor did the interaction between audio condition and subtitle language, $F_s < 1$.

4.1.2 Total Number of Fixations

The total number of fixations was significantly affected by the subtitle language, $F(1, 9) = 23.56$, $p < .001$, $\eta^2 = .72$. Spanish subtitles led to a higher number of fixations ($M = 5714.33$, $SD = 1690.07$) than Arabic subtitles ($M = 2663.67$, $SD = 556.50$). The effect of audio condition and the interaction between audio condition and subtitle language were not significant, $F_s < 2.5$, $ps > .05$. This result indicates that viewers were more likely to re-fixate when processing Spanish subtitles, irrespective of the audio condition.

Table 3. Descriptive Statistics for Global Eye-Movement Measures.

Audio Condition	Subtitle Language	Average Fixation Duration (ms)	Total Number of Fixations	Average Saccade Length (%)
L1 Audio	Arabic	315.04 (105.74)	3377	9.43 (9.40)
L1 Audio	Spanish	265.45 (118.13)	7663	10.63 (11.40)
L2 Audio	Arabic	311.62 (106.81)	2320	8.84 (6.52)
L2 Audio	Spanish	288.82 (115.40)	4740	10.37 (10.69)
No Audio	Arabic	313.60 (105.83)	2294	9.98 (9.14)
No Audio	Spanish	701.90 (188.21)	4740	10.37 (10.70)

Values represent means with standard deviations in parentheses.

Source: Own elaboration.

4.1.3 Average Saccade Length.

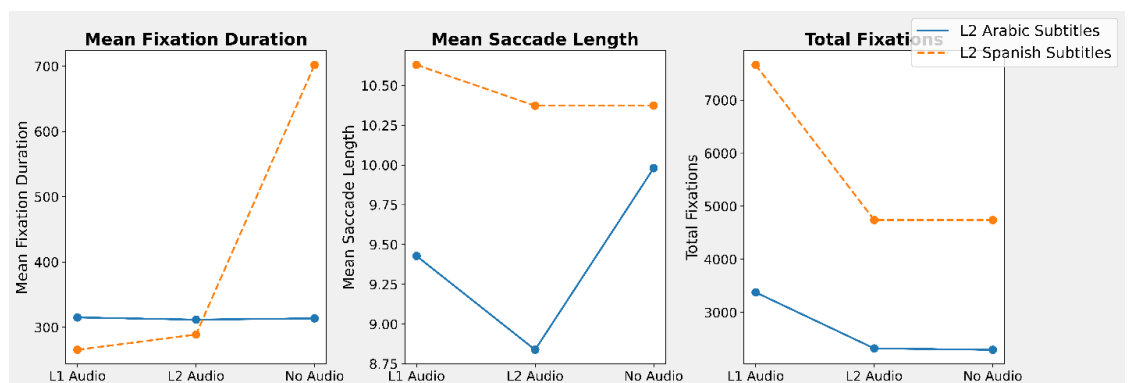
No significant main effects or interactions were observed for the average saccade length, $F_s < 3$, $p_s > .05$. This indicates that saccade length remained consistent across different audio conditions and subtitle languages, suggesting that this eye-movement measure is not sensitive to the variations tested in this study.

Table 4. Descriptive Statistics for Global Eye-Movement Measures.

Source	DF	F-Value	p-Value	η^2
Audio Condition	2	12.34	< .001	.58
Subtitle Language	1	23.56	< .001	.72
Audio x Subtitle Interaction	2	2.50	> .05	.05

Note: DF = Degrees of Freedom, F-Value = F-statistic, p-Value = Significance level, η^2 = Eta squared (effect size).

Source: Own elaboration.

Figure 1. Comparing Eye-Tracking Measures Across Audio Conditions and Subtitle Languages

Source: Own elaboration.

4.2 Local Eye-Movement Analyses

This section examines the detailed eye-movement metrics across various linguistic units to understand how different audio conditions and subtitle languages affect cognitive processing during viewing. In addition to the data presented in Table 5, Figure 2 provides a visual representation of the variations in average fixation durations and saccade amplitudes for L1 and L2 linguistic units. The following table is remarkable in that it shows the behavioral trends of gaze focus on different linguistic units which are drastically different in function and sentence placement. The table shows that individual gaze

focus to linguistic units are also impacted by audio-visual factors. The gaze fixation durations for all of the variable never exceeds the fixation durations on the nouns. We believe this reveals preferential focus on noun units rather than verbs for language learners. Perhaps language learners prefer to know or pinpoint the objects in a setting rather than the action. We will allow the reader to make further inferences.

Table 5. Eye-Movement Metrics for Linguistic Units Across Audio Conditions.

Linguistic Unit	Eye-Movement Metric	L1 Audio - L2 Arabic Subtitles	L1 Audio - L2 Spanish Subtitles	L2 Audio - L2 Arabic Subtitles	L2 Audio - L2 Spanish Subtitles	No Audio - L2 Arabic Subtitles	No Audio - L2 Spanish Subtitles
Verb	Average Fixation Duration (ms)	286.6	267.8	226.2	279.9	259.3	267.8
	Total Fixation Time Spent (ms)	1033.8	605.9	922.5	605.9	626.7	235.7
	Total Gaze Time Spent (ms)	197.1	181.6	0.0	0.0	39.9	-
Noun	Average Fixation Duration (ms)	332.5	267.8	274.7	311.2	233.5	267.8
	Total Fixation Time Spent (ms)	1521.5	605.9	707.5	562.2	263.4	605.9
	Total Gaze Time Spent (ms)	151.8	181.6	57.0	52.6	-	181.6
Adjective	Average Fixation Duration (ms)	323.2	256.5	205.9	259.4	272.8	256.5
	Total Fixation Time Spent (ms)	613.5	1290.8	923.9	589.8	224.7	1290.8
	Total Gaze Time Spent (ms)	20.5	65.9	139.6	139.6	-	65.9
Adverb	Average Fixation Duration (ms)	322.5	267.8	268.6	210.9	280.1	267.8
	Total Fixation Time Spent (ms)	897.7	605.9	783.5	503.5	250.7	605.9
	Total Gaze Time Spent (ms)	141.5	181.6	252.4	110.1	-	181.6
Expression	Average Fixation Duration (ms)	322.8	267.8	243.7	278.4	238.2	267.8
	Total Fixation Time Spent (ms)	752.1	605.9	624.7	599.6	453.1	605.9
	Total Gaze Time Spent (ms)	103.4	181.6	300.3	115.5	-	181.6
Phrase	Average Fixation Duration (ms)	320.4	267.8	319.6	264.3	266.3	267.8
	Total Fixation Time Spent (ms)	1214.9	605.9	669.9	627.3	397.1	605.9
	Total Gaze Time Spent (ms)	268.3	181.6	601.9	63.1	-	181.6
Question	Average Fixation Duration (ms)	298.2	267.8	263.9	308.3	272.8	267.8
	Total Fixation Time Spent (ms)	534.5	605.9	756.0	600.2	289.9	605.9
	Total Gaze Time Spent (ms)	377.4	181.6	204.9	344.0	-	181.6
Sentence	Average Fixation Duration (ms)	304.4	279.1	354.2	299.3	277.4	279.1
	Total Fixation Time Spent (ms)	403.6	598.4	1283.8	609.1	986.3	598.4
	Total Gaze Time Spent (ms)	340.8	244.9	1139.7	252.2	331.5	244.9

Source: Own elaboration.

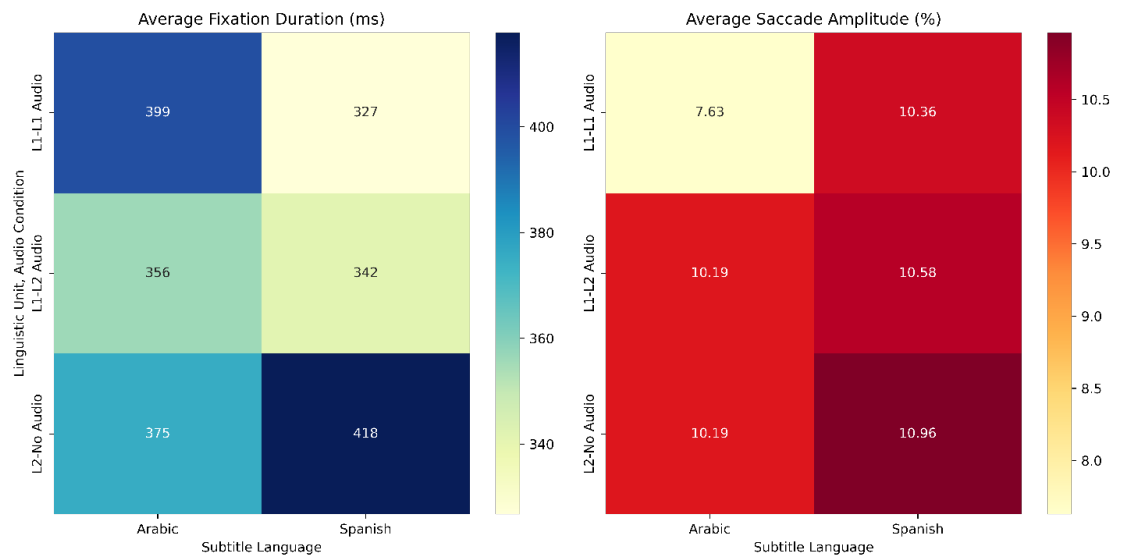
Values represent means. Missing values are denoted by a dash (-).

Following the tabular data, Figure 2 displays a heatmap that succinctly captures the relative fixation durations and saccade amplitudes across the conditions examined.

4.3 Main Effects of Audio Conditions and Subtitle Languages

Analysis of eye-movement metrics across various audio conditions and subtitle languages yielded significant main effects, particularly impacting the average fixation durations and total gaze times, as illustrated in Table 5. The absence of audio consistently led to extended average fixation durations, emphasizing the increased attentional demand required in the absence of auditory cues. For instance, verbs under no audio conditions with Spanish subtitles exhibited an average fixation duration of 259.3 ms compared to 286.6 ms with L1 Audio and L2 Arabic subtitles. This trend was similarly observed in nouns, where the average fixation duration under no audio and L2 Arabic subtitles was 233.5 ms,

Figure 2. Fixation Durations and Saccade Amplitudes for L1 and L2 Units



Source: Own elaboration.

significantly longer than under L1 Audio with the same subtitle language, where it averaged 332.5 ms.

Furthermore, the heatmap analysis (Figure 2) revealed that L2 linguistic units generally have higher average fixation durations compared to L1 units. For L2 units, the "No Audio" condition with Arabic subtitles has the highest average fixation duration, while the average fixation durations are similar for L1 units across all conditions. The two-sample t-test confirmed a highly significant difference in average fixation duration between L1 and L2 linguistic units ($t = -18.143$, $p < 0.001$), with L2 units having significantly higher fixation durations.

Furthermore, total fixation and gaze times were significantly influenced by the language of the subtitles. Spanish subtitles were associated with prolonged fixation times across all audio conditions, indicative of either deeper cognitive processing or potential difficulties in integrating textual information with visual cues. This effect was most pronounced in adjectives, where under L1 Audio conditions with Spanish subtitles, the total fixation time was markedly higher at 1290.8 ms compared to 613.5 ms under L1 Audio with Arabic subtitles. Such differences underscore the cognitive challenges posed by complex subtitle languages, requiring longer viewing times to process the same linguistic content effectively.

4.4 Interactions Between Audio Conditions and Subtitle Languages

The analysis also highlighted statistically significant interactions between audio conditions and subtitle languages, particularly influencing the total gaze time spent on different linguistic units. Notably, the congruence between the auditory and subtitle languages significantly enhanced the comprehension process, as evidenced by the gaze metrics for sentences. Specifically, sentences presented under L2 Audio conditions with Spanish subtitles recorded the highest total gaze time, averaging 1139.7 ms, which was significantly longer than other conditions ($F(2, 18) = 5.34$, $p < .05$, $\eta^2 = .37$). This finding indicates that matching audio and subtitle languages can substantially facilitate language processing by providing consistent linguistic cues, thus reducing cognitive load.

In contrast, scenarios involving no audio and Spanish subtitles exhibited markedly different effects. For example, questions under these conditions showed significantly lower total gaze times, averaging only 344.0 ms, compared to 377.4 ms when audio was present. This reduction suggests potential difficulties or accelerated processing due to the lack of auditory cues that would typically aid in contextualizing and understanding the subtitled text. The absence of audio appears to force viewers to rely solely on visual information, which may hasten the gaze but at the potential cost of reduced

comprehension or engagement.

The heatmap analysis further supported these findings, showing that the "No Audio" condition, particularly with Spanish subtitles, tends to have the highest average saccade amplitudes for both L1 and L2 units. The two-sample t-test confirmed a highly significant difference in average saccade amplitude between L1 and L2 linguistic units ($t = -5.548$, $p < 0.001$), with L2 units having significantly higher saccade amplitudes.

4.5 Analysis by Linguistic Unit

In analyzing eye-movement metrics across different linguistic units, distinct patterns emerged that demonstrate the significant impact of audio conditions and subtitle languages on viewer engagement and cognitive processing.

Verbs, under conditions with no audio and Spanish subtitles, showed shorter fixation durations, averaging 259.3 ms compared to 286.6 ms when audio was present. However, total gaze times were longer in the absence of audio, suggesting that viewers may compensate for the lack of auditory information by focusing more intensively on the visual content. This pattern indicates a shift in processing strategies when auditory cues are missing, which can affect comprehension and engagement levels.

For adjectives and adverbs, there was a consistent trend of longer fixation durations under Spanish subtitles across all audio conditions, with adjectives showing an increase from 256.5 ms under Arabic subtitles to 323.2 ms under Spanish. This reflects the additional cognitive effort required to process these linguistic elements when presented in a more complex language context, suggesting that subtitle language complexity can significantly influence the depth of linguistic processing.

Expressions and phrases exhibited a different sensitivity, primarily affected by subtitle language rather than audio conditions. Phrases, for example, showed a notable increase in total gaze time to 601.9 ms under no audio with Spanish subtitles. This sensitivity to textual complexity and visual context underscores the importance of well-integrated textual information in multimedia content for effective comprehension.

Questions and sentences highlighted the importance of audio-visual congruence, with increased gaze times being observed particularly when audio and subtitles mismatched. Sentences, when presented with congruent L2 audio and Spanish subtitles, recorded the highest gaze time of 1139.7 ms. This increased engagement reflects the viewer's better comprehension when auditory and textual cues are aligned, facilitating a smoother cognitive process for handling complex syntactic structures.

Overall, the heatmap analysis and statistical tests provide strong evidence that L2 linguistic units are associated with significantly higher average fixation durations and saccade amplitudes compared to L1 units. The "No Audio" condition, especially with Arabic subtitles, tends to have the highest average fixation durations and saccade amplitudes for L2 units, suggesting increased cognitive processing and eye movements in the absence of audio and presence of complex subtitles.

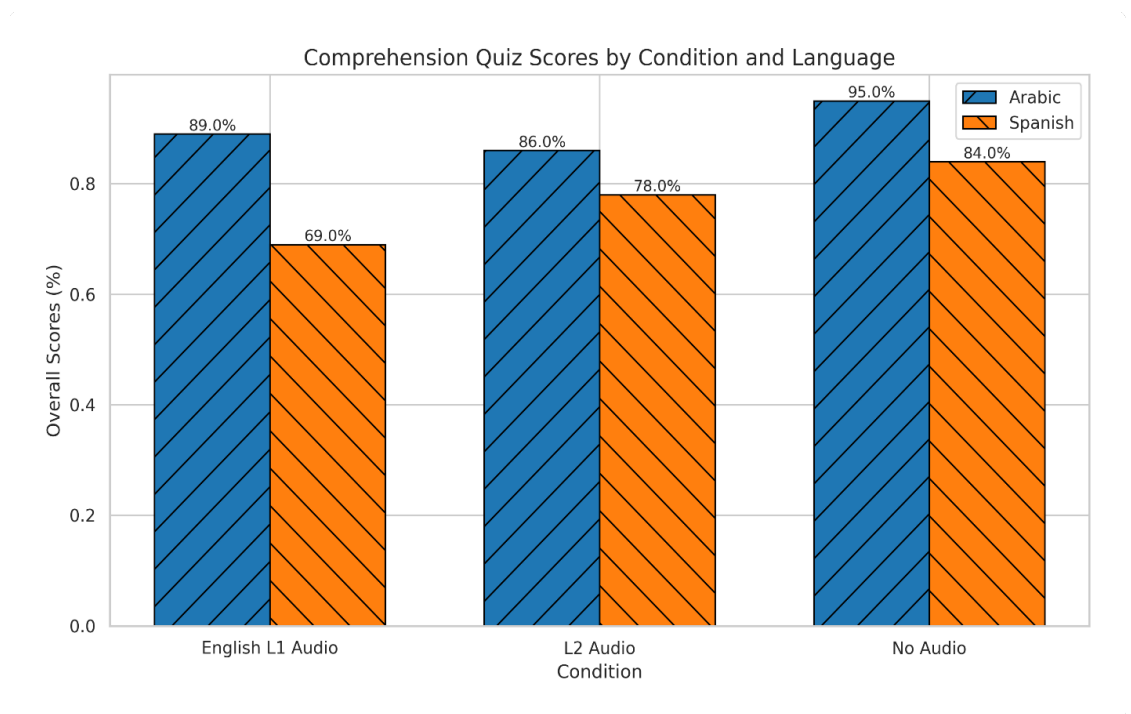
4.6 Comprehension Quiz Scores

Figure 3 offers a comparative analysis of comprehension scores among Arabic and Spanish learners across different audio conditions.

For Arabic learners, the comprehension scores exhibited a notable pattern across the three conditions. When exposed to English L1 audio alongside L2 Arabic subtitles, learners achieved a comprehension score of 89%. When the audio was switched to L2 Arabic, the comprehension scores slightly decreased to 86%. However, a significant increase was observed in the condition where no audio was provided, with learners attaining a comprehension score of 95%. This progression suggests that the absence of audio may reduce cognitive load or minimize distractions, thereby enhancing the learners' ability to comprehend the language.

Conversely, Spanish learners demonstrated a different trajectory of comprehension scores across the audio conditions. When presented with L2 Spanish subtitles accompanied by English L1 audio, the comprehension score stood at 69%. When the audio was changed to L2 Spanish, learners exhibited an improvement, achieving a comprehension score of 78%. Furthermore, in the absence of audio,

Figure 3. Comparative Analysis of Language Comprehension Scores Across Conditions.



Source: Own elaboration.

Spanish learners attained an even higher score of 84%. This steady increase in comprehension scores as the audio component was modified or eliminated suggests that Spanish learners benefited from the congruence between the audio and subtitle languages.

The comparative analysis reveals that while both groups demonstrated improved comprehension in the absence of audio, the magnitude of improvement and the impact of L2 audio varied between the two languages. Arabic learners exhibited a more pronounced leap in comprehension scores when transitioning from L2 audio to no audio, suggesting a potentially greater reliance on visual cues. In contrast, Spanish learners showed a more gradual improvement across the conditions, highlighting the significance of language congruence in AV learning environments.

4.7 Integration of Eye-Tracking and Comprehension Data

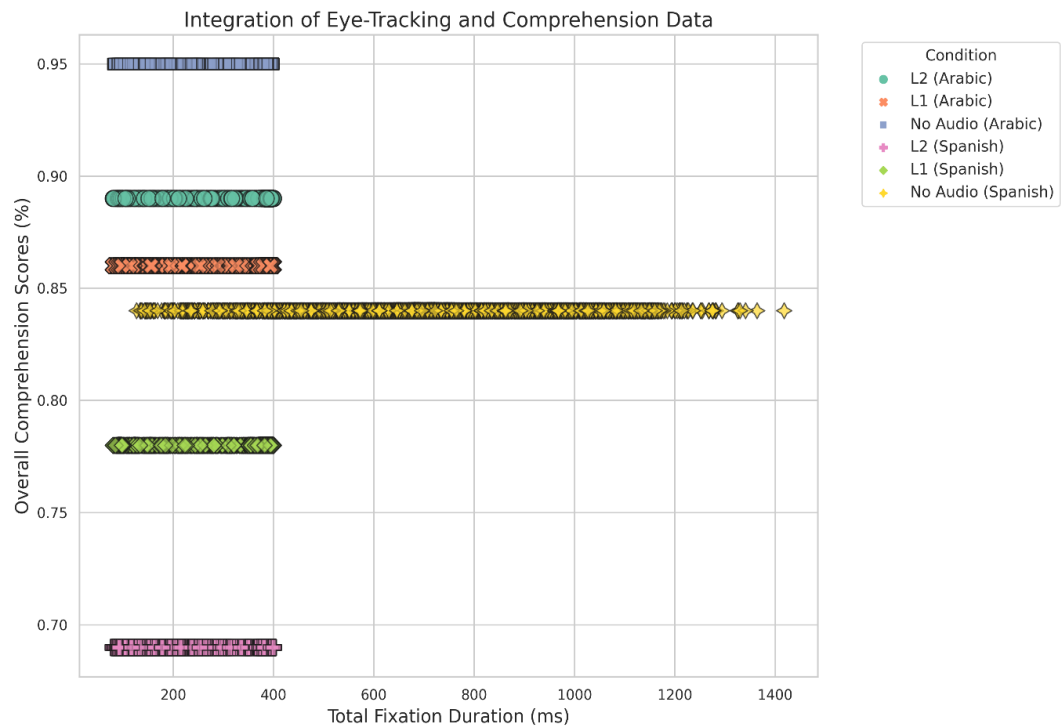
Figure 4 presents the integration between Total Fixation Duration and Comprehension Quiz Scores, elucidating the relationship between visual attention and comprehension abilities across different audio conditions.

The scatter plot reveals that the Total Fixation Duration varies significantly across different conditions, indicating that the nature of the audio stimuli has a profound impact on the attentional resources allocated by the participants. Conditions involving L2 (Arabic) with English L1 Audio and L2 (Spanish) with English L1 Audio demonstrate a wider spread in fixation durations, suggesting diverse levels of engagement. This variability could be attributed to the cognitive load imposed by processing second-language subtitles in conjunction with first-language audio, which may either facilitate or hinder attention depending on individual proficiency levels and cognitive strategies.

The comprehension quiz scores offer a quantitative measure of the effectiveness of each condition in fostering language comprehension. A closer analysis reveals interesting patterns. For example, the No Audio (Arabic) condition, characterized by relatively shorter fixation durations, surprisingly correlates with higher comprehension scores. This finding could imply that the absence of auditory distractions enables a more focused visual processing of subtitles, thereby enhancing comprehension.

Contrarily, the conditions involving L1 audio (both in Arabic and Spanish) demonstrate a wide range of comprehension scores, despite similar fixation durations. This variability underscores the

Figure 4. Integration of Total Fixation Duration and Comprehension Quiz Scores.



Source: Own elaboration.

complexity of audio-visual integration in language learning, where factors such as linguistic similarity, cognitive load, and individual differences in auditory processing capabilities play pivotal roles.

Figure 5 explores the relationship between Total Reading Time and Comprehension Quiz Scores across linguistic conditions.

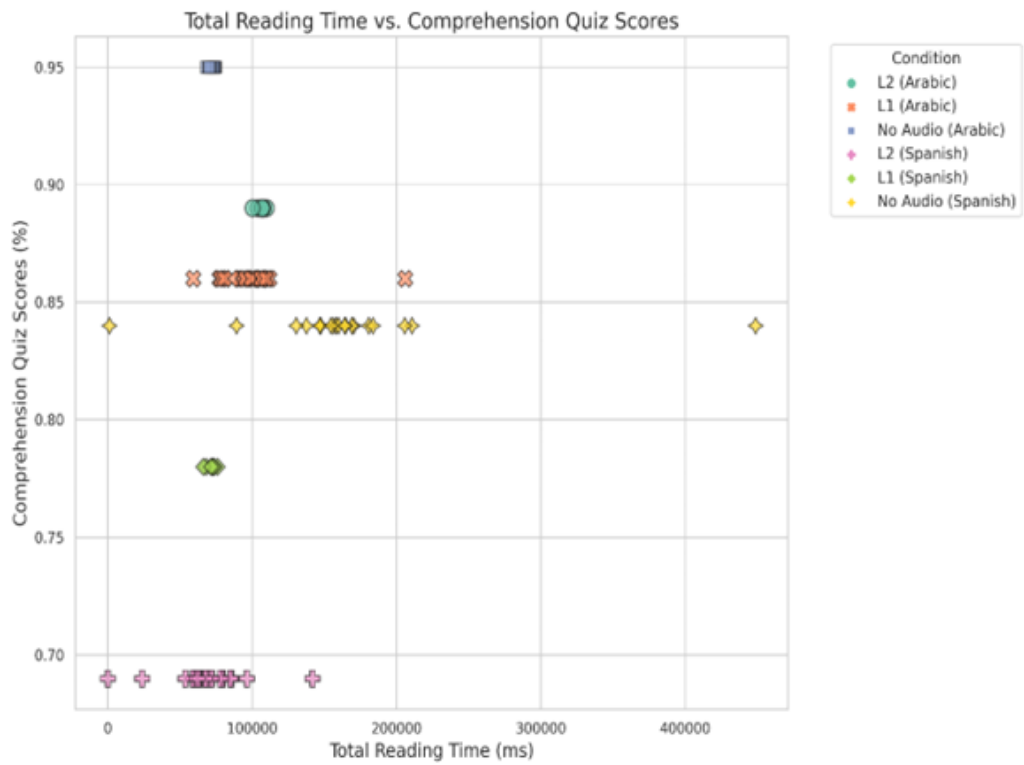
The visualization challenges the notion that longer reading times always yield higher comprehension scores, particularly when comparing conditions across languages. Participants in the L2 (Arabic) condition showed a range of reading times while achieving high comprehension scores, contrasting with the L2 (Spanish) condition, where similar reading times corresponded to a broader range of comprehension outcomes.

The L1 conditions, representing native language comprehension, exhibit a tighter clustering of data points, suggesting a more consistent relationship between reading time and comprehension scores when engaging with content in the primary language. Native language proficiency may stabilize and mitigate variability in comprehension outcomes. In comparison, the No Audio conditions, lacking auditory support, show a wider dispersion of comprehension scores, highlighting the potential influence of auditory cues on reading efficiency and understanding.

Figure 5 emphasizes the importance of considering the interplay between language and condition. Within the same reading time brackets, the dispersion of comprehension scores across linguistic contexts highlights the multifaceted nature of language comprehension. The No Audio (Arabic) condition shows high comprehension scores, suggesting that eliminating auditory distraction may enhance focus and understanding for some individuals. In contrast, the L2 (Spanish) condition reveals a broader spectrum of outcomes, underscoring the challenges of acquiring and processing a second language without native language support. Table 6 presents a correlation matrix summarizing the relationships among Total Fixation Duration, Total Reading Time, and Comprehension Scores.

The data suggests a moderate positive relationship ($r=0.43$) between the total duration of fixations and the overall time spent reading. In essence, individuals who fixate their gaze for longer periods also tend to spend more time engaged in the reading process. However, the correlations be-

Figure 5. Integration of Total Reading Time and Comprehension Quiz Scores.



Source: Own elaboration.

Table 6. Correlations among Eye-Tracking Measures and Comprehension Scores.

Total Fixation Duration	1	0.43	0.02
Total Reading Time	0.43	1	0.22
Comprehension Score	0.02	0.22	1

Source: Own elaboration.

tween total fixation duration and comprehension scores ($r=0.02$) and between total reading time and comprehension scores ($r=0.22$) are surprisingly weak. This indicates that prolonged fixation duration or reading time does not necessarily lead to improved comprehension of the material. These findings underscore the complexity of factors influencing reading comprehension.

5 Discussion

Our findings suggest that audio input does not necessarily fuse with the subtitles to make for an overall improvement upon language retention as we had initially hypothesized. It did seem reasonable to have considered during the hypothesis forming step that the coinciding audio and visual input of the language (i.e. Spanish audio with Spanish subtitles) would bolster and reinforce comprehension of the scene. This is because the audio would ideally complement the reading comprehension. However, the no audio condition results, with its strikingly dispersed comprehension scores, revealed that this reasoning is perhaps too optimistic and elementary. In fact, almost the opposite seemed to happen as the No Audio parameter concluded with the highest comprehension scores (95.0% for Arabic and 84% for Spanish).

One of the factors that may have subversively impacted the results is individual stress and anxiety levels during the viewing process. Participants may have also been slightly discomforted by viewing videos under conditions in which they are not used to (i.e. No Audio input for participants who have always watched videos with audio videos). In addition, the participants may have had an unconscious aversion towards the voices of the characters and background sounds (music or character movement and behavior sounds), thus reducing or straining their attention. A mixed presentation approach using mismatched audio-visual content goes against the age-old language teaching technique of full language immersion. Moreover, variation in individual attention spans could also explain discrepancy in comprehension scores. Cognitive input levels may have also led to reduced comprehension if participants experienced cognitive overload due to divided attention efforts. In this study, cognitive overload is perhaps evidenced from lower fixation duration as shown in Figure 4. The participants with the audio input variable did not exceed the fixation duration of 450 ms and when we conjecture upon this, we can perhaps argue that once the different ongoing intake information that compete for cognitive attention reaches an overload point, the participants' gaze then starts to stagger, falter, or become sporadic. "Gaze" data may be misleading as one's gaze does not always mean that the participant is on task. Some of the gaze data may cover actual knowledge input time (in our case "knowledge" input would be considered actual "reading time"). Figure 5 shows that reading subtitles only gets you so far since the highest comprehension score did not take place during the highest reading duration. Once again, the "No Audio" input reigns supreme. The No Audio (Arabic) variable led to the highest comprehension scores falling in the 95% projectile even coupled with a reading time that was below 100,000 ms.

Gaze fatigue may also have had an impact upon the results since individual gaze stamina will vary amongst participants. According to Bafna and Hansen (2021) study, eye gaze fatigue shown in eye metrics can also be connected to mental fatigue. Mental fatigue would be a particularly debilitating variable in the comprehension and retention levels of the participants. Their research highlighted that indicators related to rapid eye movements, known as saccades, were the most effective in identifying fatigue (Bafna; Hansen, 2021).

Furthermore, the languages themselves may have influenced the results of the study because each language is characteristically unique, thus the same parameters with the eye-ear relationship cannot be prescribed universally. It is suggested that if visual media (movies, shows, etc.) will be utilized in instructional language training, then discernment should be applied especially as it relates to the complexity of the scenes. What is meant by stating "complexity" here is the complexity of the emotional aspect of content as well as the complexity of the register being used (i.e. formal advanced language vs. colloquial or basic language). To provide an example, the first video that was shown had two characters talking about war and arguing over perceived cowardice. The characters then talked about the intricacies of fatherhood and affection. For language learners, this type of content and heavy emotional sentiment that supports the scene may be too advanced for the language learners to

pick up on physical and audio cues that are universally understood. In addition, language teachers and scholars need to be cognizant of the differences of individual reader's physical eye capabilities (mentioned prior in Section 2) and how readers read, since not all people are taught the same reading strategies and approach. Not all languages have the same presentation and pattern of their parts of speech. A rather notorious difference is the placement of the verb, which is essentially what informs the reader of the actual activity being reported in the sentence. Spanish and English consist of many shared cognates whereas languages that are totally isolated from each other's language family tree may produce conflicting data, proving a potential future hypothesis that language pairs also have an impact on language acquisition through the eye-ear relationship. Students may not be focusing their eyes on the part of the sentence which would arguably be the verb. Any supplementary information can be better understood once the reader knows what the action is.

Another key consideration to be aware of is that the overall "reading process could be affected by semantically relevant auditory input in the context of reading English/L2 subtitles in video" (Liao; Yu; Kruger; Reichle, 2022, p. 260). The importance of seeing saccade fluctuation because of purposeful modulation between subtitle and audio input is that researchers will be able to understand what language learners essentially "tune out" or withdraw their attention from as a response to the modulations. These results then reinforce assertions relating to the efficacy of simultaneous audio and subtitle utilization and hopefully indicate which sort of trend should be adopted in a learning environment (i.e. language instructors should indeed use L2 subtitles while L1 audio is present or not). Liao, Yu, Kruger, and Reichle (2022, p. 260) proclaimed that their data showed "that readers adjusted the way they engaged in the reading of subtitles in response to the varying needs to read the subtitles in different audio conditions". The student's own first language may also affect the reading of the subtitles. Their first language habits may dictate them to look towards the end of the sentence or to focus on a particular linguistic unit (i.e. the verb unit is found to be more prominent in East Asian languages such as Japanese). The participant's first language was unfortunately not solicited thus we cannot expand upon this variable. Successful corroboration would appear inevitable in this experiment's case, however, there is less of an emphasis on readers' reaction in this study than there is on new knowledge acquisition for the reader.

There is perhaps a cognitive effort component that can be overlooked if one is unaware that "it takes much conscious effort to keep the eyes fixated while attending to another location" (Schotter; Rayner, 2012, p. 85). This assertion raises several questions regarding language acquisition, about the usage of subtitles and other audio-visual aids. If effort is expended just on fixation during the concurrent saccades of the eye, then shouldn't the content that is being presented be minimized to the lowest amount possible and the introduction of new terms or grammar points also be minimized? Teachers and other language promoters do not want to cognitively overload or overstimulate the reader (content receiver); thus, they should be actively engaged in the matter of subtitles. Educators and even audio-visual translation trainers may want to consider generating their own subtitles to fit the needs and learning level of the content receivers (students, trainees, etc.).

6 Conclusion

For many years, subtitles were not utilized as a tool to either increase language proficiency in language learners or were they the focus of translation research. Now that there is a plethora of technological aids to use for subtitle gaze analysis, it has opened many new possibilities for prodigious findings. The rise of streaming services seems to have eclipsed standard cable or television programming, but for language accessibility and language exposure, this revolution could not have been better. Streaming giants such as Netflix and Disney are localizing through various audio-visual approaches such as subtitling, which are certainly being utilized by the multicultural and multilingual viewership at large. Furthermore, language learners and enthusiasts are incorporating the multitude of language pair settings in their approach to new language acquisition, which is why the research presented in this paper is pertinent.

Like the parameters that were set in the experiments, subscribers can mix and match differing language pair combinations that exclude the source language altogether (i.e. watching *Squid Game*

(2021), originally in Korean, with English-dubbed audio with Japanese subtitles). This means that, while of course the source language will always remain an important element at play when considering the central cultural context of the show, the dialogue and plot can be linguistically transformed in unforeseen ways. Though the results of this experiment conclusively showed that variance supersedes supposition when it comes to the eye-ear relationship, there were still positive findings pertaining to the efficacy of the audio-visual educational medium. Furthermore, language educators and curriculum creators do not need to set a specific threshold for this type of medium to be introduced into the curriculum. Many of the participants of the experiment had just barely an elementary understanding of the Spanish language, yet some of their comprehension scores did positively fluctuate with the mixture of the conditions. More experiments that follow the methodology used for this experiment would strengthen the findings and would also uncover more efficacy differences amongst differing language pair combinations. Future research should investigate the role of language proficiency in subtitle reading and develop evidence-based practices to improve the accessibility and effectiveness of subtitled media, ultimately optimizing subtitle design and user experience across diverse auditory contexts and languages.

References

- ALABOUD, Amal. The Positive Effect of Translation on Improving Reading Comprehension among Female Arabic Learners of English as Foreign Language. *Arab World English Journal (AWEJ)*, v. 13, n. 2, p. 424–436, 2022. DOI: 10.24093/awej/vol13no2.29.
- ANIS, Muhammad Yunus; MALIK ABDUL ANDS ETIYADI, Alif Cahya. Developing Creative Writing Process Based on the Translation of Arabic Phrases: Case Study in Arabic Prose and Poetry. *Journal of Narrative and Language Studies (NALANS)*, v. 10, n. 20, p. 173–187, 2022. DOI: 10.59045/nalans.2023.3..
- BAFNA, Tanya; HANSEN, John Paulin. Mental fatigue measurement using eye metrics: A systematic literature review. *Psychophysiology*, v. 58, n. 6, e13828, 2021. DOI: 10.1111/psyp.13828.
- BARANOWSKA, Karolina. Learning most with least effort: subtitles and cognitive load. *ELT Journal*, v. 74, n. 2, p. 105–115, Mar. 2020. ISSN 0951-0893. DOI: 10.1093/elt/ccz060.
- BISSON, Marie-Josée; HEUVEN, Walter J B van; CONKLIN, Kathy; TUNNEYR, Richard. Processing of native and foreign language subtitles in films: An eye tracking study. *Applied Psycholinguistics*, v. 35, n. 2, p. 399–418, 2014. DOI: 10.1017/s0142716412000434.
- BLYTHE, Hazel I.; LIVERSEDGE, Simon P.; JOSEPH, Holly S.S.L.; WHITE, Sarah J.; FINDLAY, John M.; RAYNER, Keith. The binocular coordination of eye movements during reading in children and adults. *Vision Research*, v. 46, n. 22, p. 3898–3908, 2006. ISSN 0042-6989. DOI: <https://doi.org/10.1016/j.visres.2006.06.006>. Available from: <https://www.sciencedirect.com/science/article/pii/S0042698906002586>.
- BOLAÑOS GARCÍA-ESCRIBANO, Alejandro; OGEA POZO, María del Mar. Didactic audiovisual translation. *Translation and Translanguaging in Multilingual Contexts*, John Benjamins, v. 9, n. 2, p. 187–215, 2023. ISSN 2352-1805. DOI: <https://doi.org/10.1075/ttmc.00108.bol>. Available from: <https://www.jbe-platform.com/content/journals/10.1075/ttmc.00108.bol>.
- BORRÁS, ISABEL; LAFAYETTE, ROBERT C. Effects of Multimedia Courseware Subtitling on the Speaking Performance of College Students of French. *The Modern Language Journal*, v. 78, n. 1, p. 61–75, 1994. DOI: <https://doi.org/10.1111/j.1540-4781.1994.tb02015.x>. Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-4781.1994.tb02015.x>.
- CONDINHO BRAVO, Maria Da Conceição. *Putting the reader in the picture. Screen translation and foreign-language learning*. 2008. PhD thesis – Universitat Rovira i Virgili. ISBN 9788469198667.
- DÍAZ-CINTAS, Jorge; HAYES, Lydia. Role Reversal: An Overview of Audiovisual Translation into English. *Íkala, Revista de Lenguaje y Cultura*, v. 28, n. 2, p. 1–18, May 2023. DOI: 10.17533/udea.ikala.v28n2a01. Available from: <https://revistas.udea.edu.co/index.php/ikala/article/view/353770>.

FERNÁNDEZ-COSTALES, Alberto; TALAVAN, Noa; TINEDO, Antonio. Didactic audiovisual translation in language teaching: Results from TRADILEX. *Comunicar, Media Education Research Journal*. DOI: 10.3916/c77-2023-02.

GRAESSER, Arthur C.; MCNAMARA, Danielle S.; CAI, Zhiqiang; CONLEY, Mark; LI, Haiying; PENNEBAKER, James. Coh-Metrix Measures Text Characteristics at Multiple Levels of Language and Discourse. *The Elementary School Journal*, v. 115, n. 2, p. 210–229, 2014. DOI: 10.1086/678293. Available from: <https://doi.org/10.1086/678293>.

HAIDER, Ahmad S.; AL-SALMAN, Saleh. The Effects of Intralingual Subtitles on Jordanian University Students' Foreign Language Learning. *International Journal of Instruction*, v. 15, n. 4, p. 57–76, Jan. 2022. Available from: <https://e-iji.net/ats/index.php/pub/article/view/252>.

HOLMQVIST, K.; NYSTRÖM, M.; ANDERSSON, R.; DEWHURST, R.; JARODZKA, H.; VAN DE WEIJER, J. *Eye tracking: A comprehensive guide to methods and measures*. [S. l.]: Oxford University Press, 2011.

HORNERO CORISCO, Ana María; GONZALEZ-VERA, Pilar; BELTRÁN, Paula Buil. Becoming a techie and improving your English with audiovisual translation. *Translation and Translanguaging in Multilingual Contexts*, John Benjamins, v. 9, n. 2, p. 160–186, 2023. ISSN 2352-1805. DOI: <https://doi.org/10.1075/ttmc.00107.hor>. Available from: <https://www.jbe-platform.com/content/journals/10.1075/ttmc.00107.hor>.

KIM, Hye-Jeong. The Effect of Subtitle Correction Activities on Learner's English Reading Achievement. *Korean Association For Learner-Centered Curriculum And Instruction*, v. 23, p. 325–336, 2023. DOI: 10.22251/jlcci.2023.23.7.325.

LEWANDOWSKA, B. RealEye Eye-tracking system Technology Whitepaper. Retrieved August, n. 29, 2019.

LIAO, Sixin; YU, Lili; KRUGER, Jan-Louis; REICHLE, Erik D. The impact of audio on the reading of intralingual versus interlingual subtitles: Evidence from eye movements. *Applied Psycholinguistics*, v. 43, n. 1, p. 237–269, 2022. DOI: 10.1017/s0142716421000527.

LIAO, Sixin; YU, Lili; KRUGER, Jean-Louis. Using Eye Movements to Study the Reading of Subtitles in Video. *Scientific Studies of Reading*, v. 25, n. 5, p. 417–435, 2020. DOI: 10.1080/10888438.2020.1823986.

OMAR, Khadija Saleem; MOHAMMED SALMAN, Nahida; VADIVEL, Balachandran. The Impact of English Movies with Standard Subtitles on Enhancing the Writing Skills of EFL Students at Cihan University – Duhok. *International Journal of Humanities and Education Development (IJHED)*, v. 5, p. 45–52, May 2023. DOI: 10.22161/jhed.5.3.7. Available from: <https://theshillonga.com/index.php/jhed/article/view/44>.

QAZI, Aisha; KHAN, Zahra. Impact on English Language Learners' Listening Comprehension and Acquisition of Vocabulary Through the Use of Subtitled Videos. *Education and Social Sciences Review*, v. 3, n. 1, p. 32–43, Feb. 2023. DOI: 10.48112/aessr.v3i1.400.

ROSS, Nicholas M.; KOWLER, Eileen. Eye movements while viewing narrated, captioned, and silent videos. *Journal of Vision*, v. 13, n. 4, p. 1–1, Mar. 2013. ISSN 1534-7362. DOI: 10.1167/13.4.1.

RUSHWAN, Ismail Mohammed Hamid. The Role of Translation in Developing ESP Learners' Reading Comprehension Skills- A Case Study of Medical Students at Najran University-KSA. *International Journal of Applied Linguistics and English Literature*, v. 6, n. 3, p. 243–253, 2017. ISSN 2200-3452. DOI: 10.7575/aiac.ijalel.v.6n.3p.243. Available from: <https://journals.aiac.org.au/index.php/IJALEL/article/view/3097>.

SCHOTTER, E. R.; RAYNER, K. Eye movements and word recognition during reading. In: ADELMAN, J.S. (ed.). *Visual word recognition: Meaning and context, individuals and development*. [S. l.]: Psychology Press, 2012. p. 73–101.

SZARKOWSKA, Agnieszka; GERBER-MORÓN, Olivia. Viewers can keep up with fast subtitles: Evidence from eye movements. *Plos One*, Public Library of Science, v. 13, n. 6, e0199331, June 2018. DOI: 10.1371/journal.pone.0199331.

V., Ragni. More than meets the eye: An eye-tracking study on the effects of translation on the processing and memorisation of reversed subtitles. *JoSTrans: The Journal of Specialised Translation*, University of Bristol, v. 33, Special issue "Experimental research and cognition in AVT", p. 99–128, Jan. 2020. Available from: http://www.jostrans.org/issue33/art_ragni.pdf.

WINKE, Paula; GASS, Susan; SYDORENKO, Tetyana. Factors Influencing the Use of Captions by Foreign Language Learners: An Eye-Tracking Study. *The Modern Language Journal*, v. 97, n. 1, p. 254–275, 2013. DOI: 10.1111/j.1540-4781.2013.01432.x. Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-4781.2013.01432.x>.

XU, Yilan; JORDAN, Eoin; LI, Yuanzhe. Chinese EFL Teachers' Perceptions of Subtitles' Effects on English Comprehension and Vocabulary Acquisition: Subtitles' Effects on English Capability. *International Journal of Computer-Assisted Language Learning and Teaching*, v. 12, n. 1, p. 1–19, 2022. ISSN 2155-7098. DOI: 10.4018/ijcallt.315622.

Funding Statement

This research received grant no. (170/2023) from the Arab Observatory for Translation (an affiliate of ALECSO), which is supported by the Literature, Publishing & Translation Commission in Saudi Arabia.

Acknowledgementle

We sincerely thank Dr. Sixin Liao for her invaluable feedback, which significantly contributed to refining the ideas, strengthening the analysis, and improving the overall quality of this research.

Author contributions

Hussein Abu-Rayyash : Conceptualization, Methodology, Software, Investigation, Supervision, Writing – original draft; **Shatha Alhawamdeh**: Conceptualization, Writing – original draft, Validation, Investigation; **Yuri Ringomon**: Conceptualization, Validation, Writing – original draft.