RESLA
Revista Española de Lingüística Aplicada

https://matrix.aesla.org.es/RESLA
ISSN 0213-2028 | e-ISSN 2254-6774

Book review: Muñoz, C., & Miralpeix, I. (Eds.). (2024). *Audiovisual input and second language learning*. John Benjamins

Injung Wi

University of Western Ontario, Ontario, Canada

In today's digital era, audiovisual input has become a prevalent source of language exposure for second language (L2) learners. This edited volume, *Audiovisual Input and Second Language Learning*, focuses on the impact of audiovisual input, especially, enhanced with onscreen text (e.g., first language [L1] subtitles, L2 captions, dual subtitles, and visually enhanced captions) on L2 acquisition, presenting nine empirical studies that were carried out as part of the Subtitles in Language Learning (SUBTill) project at the University of Barcelona. This book opens with an introduction in which the editors describe its purpose and key features, followed by nine chapters (one per empirical study), and ends with a concluding chapter in which the editors reiterate the findings, connect them to previous research, and offer insights into the role of audiovisual input and onscreen text in L2 learning.

One notable feature of this volume, as highlighted by the editors, is that it comprises studies investigating a wide range of aspects of L2 learning through audiovisual input, including vocabulary, reading skills, comprehension, L2 pragmatic awareness (e.g., English requests), pronunciation (e.g., vowel contrast: /æ/ and /ʌ/), and grammar (e.g., English subjunctive constructions). The studies also considered the roles of various learner-related variables, encompassing L2 vocabulary size, L1 and L2 reading skills, language aptitude, and age in L2 learning from viewing. This review first discusses the chapters that did not employ eye-tracking methods, organized by topic, namely vocabulary studies (Chapters 1, 5, and 9) and studies focusing on other aspects, such as reading efficacy and grammar (Chapters 2 and 8). Then, it proceeds to discuss the eye-tracking studies (Chapters 3, 4, 6, and 7).

In Chapter 1, Casulleras and Miralpeix examined how L1 subtitles and L2 captions differently affect young learners' content comprehension and vocabulary development, and how aptitude, vocabulary size, and L1/L2 reading speed moderate these effects. They found that L1 subtitles are more likely to improve content comprehension, whereas L2 captions lead to better recognition of written words. They also found that the learners' aptitude, vocabulary size, and L1 reading speed were positively correlated with content comprehension. Aptitude and vocabulary size were only positively associated with

How to cite: Wi, I. (2026). Book review: *Audiovisual input and second language learning* by C. Muñoz & I. Miralpeix (Eds.). *Revista Española de Lingüística Aplicada, 39*(1), 372-376. https://doi.org/10.58859/resla.858

Corresponding author: *iwi@uwo.ca*

AESLA
Asociación Española de Lingüística Aplicada

vocabulary learning in the L1 subtitles group. However, using only the written prompts in the recognition test may have affected the test outcomes in favor of the captioned group due to the test-input modality congruency effects (e.g., Jelani & Boers, 2018; Sydorenko, 2010).

Popova and Miralpeix (Chapter 5) demonstrate the facilitative effects of repeated viewing on content comprehension and vocabulary learning. They compared repeated viewing (watching the same video twice) with viewing using *Language Reactor*, a language-learning app that enables self-regulated viewing with access to bilingual subtitles, pausing, replaying, or adjusting playback speed. The results showed no significant differences in comprehension or vocabulary gains between the two conditions, suggesting that ordinary repeated viewing with captions can be equally effective as self-regulated viewing for supporting comprehension and vocabulary learning. The qualitative data indicated that the first viewing supported content comprehension, enabling learners to better attend to linguistic features during the second viewing. Similar findings were reported by Wi and Boers (2024), who showed that repeated viewing is a common strategy used by advanced L2 learners to support understanding and vocabulary acquisition.

A study by Miralpeix, Gesa and Suárez (Chapter 9) investigated the extent to which audiovisual input can promote novice learners' vocabulary learning at First Exposure (FE). Catalan/Spanish native speakers who are fluent in English watched an English TV advertisement twice with Polish subtitles, which is a language they are not familiar with (reversed subtitling). The participants were explicitly instructed to learn the Polish words as much as possible. They found that learning at the level of meaning recognition could occur at FE through reversed subtitling. The researchers acknowledged that the observed learning gains could also be attributed to the intentional learning condition and repeated viewing.

While the previous studies have investigated vocabulary learning, Avello and Muñoz (Chapter 2) examined whether watching captioned videos could improve young learners' L1 and L2 reading efficacy, including silent reading speed and reading comprehension. The participants were primary students learning English in Chile. They revealed that watching captioned videos enhanced their L2 reading speed and text comprehension, with vocabulary knowledge and pretest scores moderating this effect. Age did not predict L2 reading skills. Interestingly, L1 reading efficacy was also improved through captioned videos despite the exposure to only L2 captions. The researchers suggest that L1 and L2 reading skills seem to complement one another during the early stages of development. The findings shed light on the potential of captions in promoting young learners' reading skills in both languages.

In Chapter 8, Pattermore and Suárez describe two studies that explored the role of language aptitude (grammatical inference ability) in L2 grammar learning through captioned and uncaptioned videos. They found a significant interaction between viewing condition (captioned vs. uncaptioned) and aptitude (lower vs. higher), indicating that aptitude had a greater effect when participants watched an uncaptioned video. In contrast, its effect was less pronounced when the video was captioned. Another finding was that learners with lower proficiency levels tend to rely more on their aptitude than those with higher

proficiency levels. These findings corroborate that captions can serve as a scaffolding tool, especially for those with lower language-learning aptitude and proficiency levels. However, it should be acknowledged that these studies have limitations due to their small sample size.

Another noteworthy feature of this volume is that many studies adopted a triangulation approach, incorporating eye-tracking technology, questionnaires, and retrospective interviews. Notably, four studies (Chapters 3, 4, 6, and 7) employed eye-tracking methods.

Finger-Boua and Muñoz (Chapter 3) compared the effects of visually enhanced and regular captions on young learners' attention and vocabulary acquisition using eye-tracking and vocabulary tests. The results indicated that learners exposed to enhanced captions outperformed those in the regular captions group on both form recall and recognition tasks. Eye-tracking data revealed longer fixation durations for the enhanced captions group at the initial stage (episode 3), although this difference diminished over time (episode 11). While methodologically original, given the scarcity of longitudinal eye-tracking studies with young learners, the use of identical target words for both tests raises potential concerns about testing effects.

Pusadas and Puimège (Chapter 4) used eye-tracking to examine beginner learners' attention during L1-subtitled and L2-captioned video viewing and the effects of pre-teaching vocabulary on comprehension. The L1-subtitled group focused more on target vocabulary and comprehension-related text and outperformed the captioned group on comprehension, suggesting beginner learners depend more on L1 subtitles for understanding. Pre-viewing instruction affected attention only in the captioned group, resulting in fewer fixations on target words and suggesting faster processing due to prior exposure. However, pre-viewing words did not significantly influence comprehension. As acknowledged by the researchers, however, the study was limited by the small number of target words and comprehension questions.

Barón, Celaya, and Martínez-Flor (Chapter 6) explored whether captions (captioned vs. uncaptioned videos) facilitate the development of L2 pragmatic knowledge (awareness of English requests) using selected scenes from *Emily in Paris*. Employing a triangulated approach with eye-tracking and retrospective interviews, they found that the captioned group outperformed the uncaptioned group on the post-test. The interview data further support this result. Eye-tracking measures of visits to the eyes and mouth revealed no significant group differences across the three clips, except for the second clip. The findings offer valuable insights into the potential of captioned videos for pragmatic development, although the multiple-choice test format used to assess pragmatic knowledge may pose issues such as guessing and input-test modality congruency effects (e.g., Jelani & Boers, 2018; Sydorenko, 2010). Given the limited research on L2 pragmatic learning through audiovisual input, this study highlights the need for further investigation.

In Chapter 7, Mora and Fouz-González examined whether using enhanced captions in two contrasting colors versus a single color and using standard orthography and IPA symbols would differentially affect Spanish learners' perceptual learning of the English

vowel contrast (/æ/–/ʌ/). Eye-tracking measures were employed to assess the effectiveness of the visual enhancement in directing learners' attention to the target sounds. The results revealed evidence of perceptual learning from viewing with enhanced captions, although the effects were not consistently observed across tasks and conditions. The eye-tracking data indicated that enhanced captions drew more attention, as reflected by longer relative fixation times compared to regular captions. The learning gains were modest, but the researchers remained optimistic about the contrastive input enhancement method, as a single intervention appeared to draw learners' attention to challenging sounds and promote their perceptual sensitivity.

The last notable feature of this volume is its inclusion of longitudinal studies with young learners, specifically primary school students aged 9 to 12 years (Chapters 1, 2, and 3). For instance, in Chapter 1, 92 Spanish/Catalan bilingual EFL learners with an average age of 11.55 years watched 20 episodes of a children's TV series once a week over five months. Similarly, the studies in Chapters 2 and 3 also adopted multi-treatment sessions where young EFL learners watched a total of 11 episodes of a TV series. Given that the impact of audiovisual input on young learners' L2 acquisition over time remains an underexplored area (e.g., Montero Perez, 2022), these chapters deepen insights into how young learners process captioned videos and benefit from prolonged exposure to them in language learning.

This edited volume brings together a series of studies that coherently examine the effects of audiovisual input on various aspects of L2 learning. Each chapter is well-supported with extensive references, and most include appendices detailing test instruments, target vocabulary, and input materials, which enhances the transparency and replicability of the research. However, a few limitations were identified during the review process. One limitation of the volume is its exclusive focus on English as a foreign language (EFL) contexts, as most studies were conducted in Spain with participants who were Spanish learners of English. Consequently, the generalizability of the findings to other linguistic and educational settings may be limited. Another limitation is that most studies in this volume employed a pretest-posttest design, leaving the long-term effects of the interventions unaddressed. Although some researchers acknowledged this limitation, incorporating a delayed posttest in future studies would be essential to evaluate the retention of learning gains over time. Lastly, the discussion could have been more informative if some of the previously mentioned methodological issues related to test modality and format (e.g., input-test modality congruency) had been more thoroughly addressed. This consideration is important given the challenges researchers face in selecting and designing appropriate and valid measures of L2 learning, particularly in studies comparing the effectiveness of captioned or uncaptioned videos or comparing various types of onscreen text (e.g., Kurokawa et al., 2024; Wi & Boers, 2025).

In conclusion, despite these limitations, this volume significantly advances second language acquisition (SLA) research on audiovisual input with onscreen text. Firstly, its breadth of focus aids in understanding the comprehensive benefits of multimodal input for L2 acquisition. Secondly, the volume demonstrates strong methodological rigor, with many

studies employing a triangulation approach that combines eye-tracking, questionnaires, and retrospective interviews. This multi-method strategy enhances the validity of findings and provides a deeper understanding of learners' processing of multimodal input. Thirdly, the volume addresses an underexplored area by including longitudinal studies with young learners, offering insights into how exposure to captioned audiovisual input over time supports language development among primary school students. This is particularly relevant in today's digital learning environments, where children increasingly encounter language through media.

I highly recommend this volume to language teachers, as many chapters offer practical pedagogical implications for using audiovisual input and onscreen text for L2 learning. It is also well-suited for graduate students and researchers interested in this area of inquiry. In particular, the volume is valuable for those looking to employ diverse research instruments, especially for integrating eye-tracking methodology into studies of second language acquisition through audiovisual input with onscreen text.

Conflict of interest and copyright / *Conflicto de intereses y derechos de autoría*
Conflict of interest / *Conflicto de intereses:* The author has no conflict of interest to declare / *El autor declara la ausencia de conflicto de intereses.*

Editorial history / *Fechas del proceso editorial*
Received / *Recibido:* 15/07/2025
Accepted / *Aceptado:* 07/11/2025
Published / *Publicado:* 01/04/2026

References

Jelani, N.A.M., & Boers, F. (2018). Examining incidental vocabulary acquisition from captioned video. *ITL International Journal of Applied Linguistics*, *169*(1), 169–190. https://doi.org/10.1075/itl.00011.jel

Kurokawa, S., Hein, A.M., & Uchihara, T. (2024). Incidental vocabulary acquisition through captioned viewing: A meta-analysis. *Language Learning*, 1–49 https://doi.org/10.1111/lang.12697

Montero Perez, M. (2022). Second or foreign language learning through watching audio-visual input and the role of on-screen text. *Language Teaching*, *55*(2), 163–192. https://doi.org/10.1017/S0261444821000501

Sydorenko, T. (2010). Modality of input and vocabulary acquisition. *Language Learning & Technology*, *14*(2), 50–73. https://doi.org/10.64152/10125/44214

Wi, I., & Boers, F. (2024). Sequential use of L1 and L2 captions: Exploring the benefits for vocabulary acquisition. *TESOL Quarterly*, *58*(1), 511–521. https://doi.org/10.1002/tesq.3243

Wi, I., & Boers, F. (2025). A synthesis of research on L2 vocabulary learning through audiovisual input and on-screen text. *Digital Applied Linguistics*, *3*, 102774. https://doi.org/10.29140/dal.v3.102774