International Journal of TESOL Studies (2026)
Vol. 8 (3) 67-81 https://doi.org/10.58304/ijts.250907

Article

Incidental vocabulary Learning from TV: Subtitles and Imagery

Kadir Kaderoğlu
University of Barcelona, Spain

Received: 12 August 2025 / Received in revised form: 6 September 2025 / Accepted: 6 September 2025 / Available online: 12 September 2025

Abstract
Incidental vocabulary learning from audiovisual input is a key area of interest, but the effects of multimedia elements like subtitles and imagery remain complex. This study investigated how on-screen text (L1 subtitles vs. L2 captions) and visual imagery influenced incidental vocabulary learning from a documentary viewing where target words occurred only once. Using a pretest-posttest control group design, it examined the effects of on-screen imagery, subtitle type (L1 vs. L2), and other moderating variables such as cognateness and prior vocabulary knowledge. Learning was assessed via form recognition, meaning recall, and meaning recognition tests. Data were analyzed using Generalized Estimating Equations to account for learner- and word-related factors. Results showed that L2 captions were uniquely effective for meaning recall. Contrary to the literature, on-screen imagery did not provide a consistent benefit and had a significant negative effect on form recognition. The findings demonstrate that prior vocabulary knowledge and cognateness remain powerful predictors of incidental vocabulary learning.

Keywords
Incidental vocabulary learning, imagery, visual support, subtitles, captions

1 Incidental Vocabulary Learning from Audiovisual Input with On-screen Text (L1 or L2 Subtitles)

Television viewing remains to be the most consumed form of media among European citizens, with 81% of the population watching TV daily, while reading printed materials has long been in decline (European Commission, 2022). Moreover, research into the media consumption habits of young foreign language learners consistently reveals a strong preference for audiovisual content over printed materials. Studies show that subtitled (L1 or L2) and non-subtitled videos are among the most frequently consumed media, and a positive correlation is repeatedly found between the extent of this consumption and learners' L2 proficiency (e.g., De Wilde & Eyckmans, 2017; Peters, 2018; Uztosun & Sundqvist, 2025).

Compelling evidence for incidental learning through audiovisual input is provided by large-scale correlational studies. De Wilde and her colleagues (2020), for example, investigated the English

Email: kkaderoglu@ub.edu

proficiency of 780 Belgian 10- to 12-year-olds who had no prior formal English instruction. They found that 10% and 14% of the children already possessed A2-level listening and speaking skills, respectively, which correlated significantly with their exposure to English-language television. Similar results have been reported in other studies as well (Gyllstad et al., 2025; Leona et al., 2021).

The role of L1 subtitles for L2 vocabulary learning may be particularly revealing considering the standard media use is subtitling, rather than dubbing, in the countries where these studies took place: the Flanders region of Belgium, the Netherlands, and Norway. Indeed, early research demonstrated that young Dutch learners acquired significant vocabulary from watching L1-subtitled documentaries (Koolstra & Beentjes, 1999). However, the facilitative effect of L1 subtitles has been contested in more recent experimental studies. Fievez et al. (2020), for instance, found no significant difference in learning gains between an L1 subtitling group and a no-treatment control group. Similarly, Peters (2019) reported no significant advantage for L1 subtitles over a no-subtitle condition at the level of meaning recall. These conflicting findings indicate that the benefits of L1 subtitles are not fully understood and warrant further investigation.

Nevertheless, these findings underscore the potential for language learning through meaning-focused activities like viewing audiovisual input with on-screen text. Thus, incidental vocabulary learning, defined as the learning of words as a by-product of a meaning-focused activity (Webb, 2020), can occur from viewing audiovisual input. This holds true regardless of whether the video material is as short as 12 minutes (Peters, 2019), as long as an hour (Ahrabi Fakhr et al., 2021; Peters & Webb, 2018) or distributed over multiple sessions with approximately 20 minutes of viewing per session (Fievez et al., 2020; Pujadas & Muñoz, 2023).

A recent meta-analysis by Webb et al. (2023) found that learning rates from viewing were lower than those from reading, listening, or reading-while-listening. However, it is likely that the low learning rates were due to the fact that the studies included in the meta-analysis did not use any textual on-screen support like L1 subtitles or L2 subtitles (captions), as both L1 and L2 subtitles have been shown to substantially boost vocabulary learning (Reynolds et al., 2022). Specifically, captioned viewing leads to greater incidental vocabulary gains than non-captioned viewing (Kurokawa et al., 2024). Indeed, studies that directly compare learning modes have found that captioned viewing consistently results in significantly higher gains than reading, listening, or reading-while-listening (Teng, 2024, 2025a, 2025b). These results indicate that the potential of audiovisual input with on-screen text is clear. Nonetheless, learning outcomes are moderated by several variables (Kurokawa et al., 2024; Reynolds et al., 2022), some of which are discussed next.

2 Factors Affecting Incidental Vocabulary Learning from Audiovisual Input

2.1 Prior vocabulary knowledge

Prior vocabulary knowledge, typically measured via receptive vocabulary size tests (e.g., Webb et al.'s (2017) the Updated Vocabulary Levels Test), is a key predictor of learning success. Research consistently reveals a Matthew effect in the relationship between prior vocabulary knowledge and incidental learning from audiovisual input, whereby learners with greater vocabulary knowledge benefit more than those with less (Ahrabi Fakhr et al., 2021; Fievez et al., 2020; Peters & Webb, 2018; Webb, 2020) This effect possibly stems from the fact that learners with greater prior vocabulary knowledge have more mental resources that they can allot to dealing with unknown vocabulary compared to those with less prior vocabulary knowledge (Webb et al., 2023). More specifically, studies have shown that each additional correct response on a prior vocabulary knowledge test increased the chances of correctly recalling a target item's meaning by 2% (Ahrabi Fakhr et al., 2021) and 3% (Peters & Webb, 2018).

2.2 Cognates

Cognateness refers to the semantic, etymological, phonological, and orthographic similarity of words between two languages (Peters & Webb, 2018). Due to this resemblance, cognates exhibit a high degree of saliency and are typically easier to notice and learn than non-cognates (Fievez et al., 2020; Peters, 2019; Peters & Webb, 2018). Further evidence is also found in De Wilde et al. (2020), where participants without prior English instruction scored much higher on Dutch-English cognates than on non-cognates in a receptive vocabulary test.

2.3 Imagery and audiovisual input

Following Rodgers' (2018) seminal work, there has been a marked increase in empirical studies on the effect of imagery (i.e., visual support) in audiovisual input (e.g., Ahrabi Fakhr et al., 2021; Fievez et al., 2020; Peters, 2019; Pujadas & Muñoz, 2023). Traces of imagery effects in audiovisual input studies, however, can be found in earlier research. Sydorenko (2010), for example, found that words with visual support yielded higher post-test scores than those without, and noted that the majority of the participants reported using visuals to infer the meanings of unknown words.

Rodgers (2018) introduced a measurement of imagery by quantitatively and systematically investigating the extent of imagery present in documentaries and narrative TV shows. The study found that audiovisual co-occurrences were significantly more frequent in documentaries, leading the author to suggest that documentaries may be more conducive to vocabulary learning due to their visual richness. Putting this claim to the test, Peters (2019) used an excerpt from the same documentary Rodgers (2018) analyzed and found that visually supported words were three times more likely to be learned than words without visual support, both at the form-recognition and meaning-recall levels. Subsequent studies have confirmed the facilitating effects of imagery (Ahrabi Fakhr et al., 2021; Fievez et al., 2020), with a recent study suggesting that duration of on-screen imagery correlates with better recall (Pujadas & Muñoz, 2023). Overall, these findings suggest that both the presence and the temporal quality of visual support are contributing factors to learning.

3 Theoretical Framework

While the positive effects of captioned audiovisual input are well-documented by cross-sectional (e.g. Peters & Webb, 2018), longitudinal (e.g. Fievez et al., 2020) and meta-analytic (e.g. Kurokawa et al., 2024) studies, it is crucial to explain the cognitive mechanisms that facilitate this learning. Two cognitive theories provide a robust foundation for understanding these processes: Dual Coding Theory (DCT) (Sadoski & Paivio, 2013) and the Cognitive Theory of Multimedia Learning (CTML) (Mayer, 2021). Together, they explain how the combination of sound, images, and text can enhance incidental vocabulary learning.

DCT postulates that human cognition operates via two distinct but interconnected systems: a verbal system for processing linguistic information and a non-verbal system for processing imagery and other sensory data (Sadoski & Paivio, 2013). Information encoded in both systems simultaneously creates multiple retrieval pathways, making memory traces more resilient and easier to access. Building on this foundation, CTML offers a model tailored to instructional contexts, arguing that "people learn more deeply from words and pictures than from words alone" (Mayer, 2021, p. 57).

When applied to the context of viewing captioned video, DCT and CTML may explain the mechanisms involved in incidental vocabulary learning. The verbal system is engaged by both the aural input (spoken narration) and the textual input (subtitles), which provide orthographic (L2 subtitles) or semantic (L1 subtitles) reinforcement for the spoken words. This textual support can alleviate cognitive load, freeing up resources for learners to segment the speech stream and notice the form and meaning of

unknown words (Peters, 2019). Simultaneously, the non-verbal system is activated by the documentary's visuals. When congruent imagery representing an unknown word appears on screen, it synchronizes with the auditory and textual input. This potentially creates an optimal learning environment by permitting a direct, powerful link between a word's form (presented aurally and textually) and its meaning (presented pictorially).

4 Rationale and Research Questions

Despite the insights provided by previous research, key methodological limitations remain in the literature. Previous research has often treated visual support as a dichotomous variable while allowing word frequency to vary, making it difficult to disentangle the effect of a visual cue from the cumulative effect of multiple exposures (e.g. Fievez et al., 2020; Peters, 2019). For example, target items occurred between 1 and 25 times in Fievez et al. (2020), but no information was provided as to whether an item with 25 occurrences had visual support for every instance, for just a few, or for only once. Furthermore, some studies have focused exclusively on concrete nouns (e.g. Pujadas & Muñoz, 2023) and have not included control groups to account for pre-testing effects (Peters, 2019). The present study was designed to address these gaps. By using only single-occurrence items of various parts of speech and incorporating a no-treatment control group, this research aims to isolate the effects of visual support and subtitling condition on incidental vocabulary learning. Consequently, the following research questions were explored:

1. Is there an effect of viewing L1 and L2 subtitled audiovisual input on the incidental learning of vocabulary items that appear only once?
2. Is there an effect of visual support on the likelihood of incidental learning of vocabulary items that appear only once in audiovisual input?
3. Is there an interaction effect of visual support and subtitling condition on the incidental learning of vocabulary items that appear only once in audiovisual input?

5 Methodology

5.1 Study design

This study employed a quasi-experimental, pre-test-post-test, between-participants design. The design was quasi-experimental because two intact classes, rather than individual participants, were randomly assigned to the treatment conditions. A control group (n = 12) was also established by randomly selecting six participants from each class. This group completed the tests without receiving any treatment, which is critical for ensuring that any observed significant learning gains resulted from the intervention itself, rather than from pre-test exposure (Nation & Webb, 2011).

5.2 Participants

Fifty-seven Catalan-Spanish bilingual students (aged 15–16) from two intact tenth-grade classes at a semi-private school comprised the initial sample after returning parental consent forms. One class was randomly assigned to the English (L2) subtitles condition (n = 21) and the other to the Spanish (L1) subtitles condition (n = 24). Since seven participants missed the post-test, the final sample consisted of 50 participants: 17 in the L2 subtitles condition, 21 in the L1 subtitles condition, and 12 participants in the control group. According to their English teacher, the overall proficiency level of both classes was approximately B1 on the CEFR scale.

5.3 Documentary input

The audiovisual input was the first 34 minutes and 15 seconds of the "Woodland" episode from the docuseries *Wild Isles* (Howard, 2023). A documentary excerpt was selected because documentaries are visually rich (Rodgers, 2018), can reduce reliance on lexical knowledge (Durbahn et al., 2024) and support incidental vocabulary learning more effectively than other audiovisual genres (Kurokawa et al., 2024). At the end of the session, all participants confirmed they had not previously seen the documentary, mitigating the potential influence of prior viewing on their behaviour.

5.4 Target items

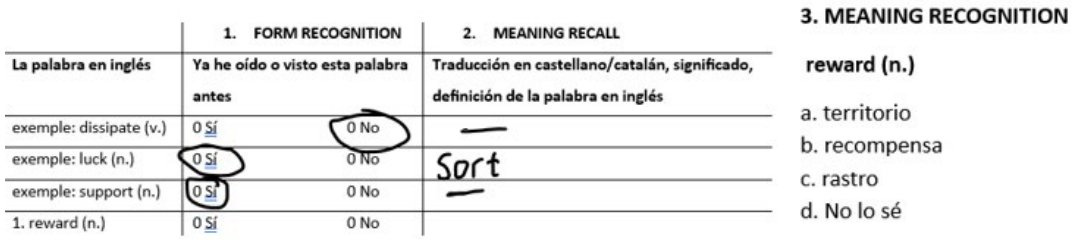
To adequately capture incidental learning gains, 32 words were selected for the 34-minute viewing period, following the rationale of Peters and Webb (2018). Furthermore, in line with Peters' (2019) inclusive selection approach, this study chose items ranging from the first 1,000-word family level (e.g., *messy*) to the tenth 1,000-word family level (e.g., *weasel*). The study included verbs and adjectives in addition to concrete nouns, as well as cognates alongside non-cognates to enhance the ecological validity. Table 1 provides a comprehensive list of the target items.

All items appeared only once in the input, with half (n = 16) classified as having visual support. Following Rodgers (2018), an item was coded as visually supported if its corresponding image was on-screen within five seconds before or after its aural or written occurrence. For example, the target item weasel was coded as visually supported because the animal appeared on-screen within five seconds of the aural/written occurrence of the word in the narration. The coding of visual support achieved 97% inter-rater reliability with a second rater. An independent-samples t-test confirmed no significant difference in concreteness ratings between the visually supported (M = 3.95) and non-supported items (M = 3.69), $t(30) = -1.01$, $p = .32$, ensuring that concreteness did not confound the results (Nation & Webb, 2011). To enhance ecological validity, the study included verbs and adjectives in addition to concrete nouns, as well as cognates alongside non-cognates. Finally, two secondary school English teachers independently rated the familiarity of all target items on a 5-point scale (1 = *definitely unknown*; 5 = *definitely known*) for average tenth grade students, reaching 88% agreement. Based on these procedures, all 32 selected items were retained for the study.

5.5 Vocabulary tests

The use of multiple measures is recommended when assessing vocabulary knowledge because knowing a word is a complex phenomenon and some tests may not be sensitive enough to capture the small increments in the word knowledge (Nation & Webb, 2011). Thus, vocabulary knowledge was assessed using three tests with good reliability (all Cronbach's $\alpha > .85$). The form recognition test required participants to indicate if they had seen or heard an item (see Figure 1 for a sample item). The meaning recall test required participants to provide an L1 translation or L2 synonym/definition (see Figure 1 for a sample item). The meaning recognition test was a multiple-choice task where participants selected the correct L1 translation from four options – correct option, two distractors, sourced from the same or a higher frequency band as the target item (Nation & Webb, 2011), and an "I don't know" option (see Figure 1 for a sample item). To control for guessing and minimize overestimation, the form recognition test included eight nonwords (Peters, 2019). Additionally, to reduce potential memory effects from pre-testing and increase test motivation, 16 relatively easy non-target items from the documentary were added to the tests. Furthermore, to avoid any modality bias favouring a specific viewing condition, a native English speaker read all test items aloud twice (Peters, 2019; Sydorenko, 2010).

Figure 1
Sample Test Items for Each Vocabulary Measure



	1. FORM RECOGNITION		2. MEANING RECALL
La palabra en inglés	Ya he oído o visto esta palabra antes		Traducción en castellano/catalán, significado, definición de la palabra en inglés
exemple: dissipate (v.)	0 Sí	0 No	—
exemple: luck (n.)	0 Sí	0 No	Sort
exemple: support (n.)	0 Sí	0 No	—
1. reward (n.)	0 Sí	0 No	

Table 1
List of Target Items

Item	Visual Support	Concreteness	Cognateness	Part of Speech	BNC/COCA frequency list
messy	1	3.59	0	adj.	1
drag	1	4.17	0	v.	2
release	1	3.24	0	v.	2
worm	1	4.9	0	n.	2
drill	1	4.4	0	v.	3
cemetery	1	4.93	1	n.	4
forehead	1	4.9	0	n.	4
peck	1	4.16	1	n.	4
protrude	1	3.34	0	v.	6
hibernate	1	2.92	1	v.	7
tattered	1	3.7	0	adj.	7
excrete	1	2.89	1	v.	8
squirt	1	3.93	0	v.	8
grub	1	3.86	0	n.	9
speckled	1	3.59	0	adj.	9
weasel	1	4.74	0	n.	10
stomach	0	4.89	1	n.	2
leap	0	3.86	0	v.	2
defeat	0	2.96	1	n.	3
grip	0	4.3	0	n.	3
reward	0	3.41	0	n.	3
damp	0	3.81	0	adj.	3
obstacle	0	3.48	1	n.	4
scent	0	3.97	0	n.	4
slim	0	3.93	0	adj.	4
sting	0	4.41	0	n.	4
roam	0	2.85	0	v.	5
sneak	0	2.82	0	v.	5
humid	0	3.48	1	adj.	6
lack	0	2.04	0	v.	6
lynx	0	4.5	1	n.	13
hermaphrodite	0	4.42	1	n.	13

Note. Each frequency band consists of 1,000 word families.

5.6 Vocabulary size test

To measure participants' receptive vocabulary knowledge, the first three 1,000-word levels of the Updated Vocabulary Levels List – UVLT – (Webb et al., 2017) were administered. This was deemed necessary due to time constraints.

5.7 Procedure

Data were collected over three sessions. In the first session, consent forms were distributed and collected the week after. During the second session, participants completed the vocabulary pre-tests and the UVLT. To mitigate order effects, the form recognition and meaning recall tests were always administered before the meaning recognition test (Nation & Webb, 2011). One week later, the treatment groups watched the documentary under their assigned condition while the control group engaged in unrelated activities. Immediately following the viewing, all participants completed a comprehension questionnaire and the post-tests. The study concluded with a full debriefing.

5.8 Scoring and analyses

All vocabulary tests were scored dichotomously (1 = *correct*, 0 = *incorrect*). To answer the research questions, a series of repeated measures logistic regressions was conducted using the Generalized Estimating Equations (GEE) procedure. This approach was selected as it accommodates the binary outcome data, accounts for the clustering of items within participants, and allows for the simultaneous inclusion of both item- and participant-level predictors in a single model (Peters & Webb, 2018).

Prior to the analyses, key statistical assumptions were verified; multicollinearity was not a concern (all VIF < 5), and the observations-per-predictor ratio was sufficient (Peters, 2019). Based on previous research (Peters, 2019), the GEE models also included cognateness, prior vocabulary knowledge, and concreteness as predictors. Similar to previous research (e.g., Fievez et al., 2020; Peters, 2019), all the predictors were fitted into the initial model first. Following a stepwise backward elimination approach, non-significant predictors were eliminated from the model one by one, and the analysis was run again. This was repeated until only the significant predictors remained in the model.

6 Results

6.1 Vocabulary size test

An initial screening of the UVLT scores revealed that some participants may have lacked the minimum proficiency required for the task. Thus, seven participants who scored 14 or lower on either the first 1000 or the second 1000 levels of the UVLT were excluded from all subsequent analyses. The final sample consisted of 43 participants.

Table 2
UVLT Vocabulary Levels Scores

Condition	1k score (max = 30)	2k score (max = 30)	3k score (max = 30)	Total score (3k) (max = 90)
Control (n = 10)	28.2 (1.99)	24.2 (4.52)	20.0 (3.83)	72.4 (8.63)
EngSub (n = 15)	28.1 (2.42)	24.5 (4.91)	22.9 (5.66)	75.6 (12.2)
SpaSub (n = 18)	28.8 (1.63)	24.3 (4.07)	22.4 (4.48)	75.6 (8.96)

Table 2 presents the final sample's vocabulary size scores. A one-way ANOVA confirmed that there were no statistically significant differences in prior vocabulary knowledge between the groups $F(2, 40) = 0.38$, $p = .690$.

6.2 Effects of L1 and L2 subtitles on incidental vocabulary learning from audiovisual input

To address the first research question, a GEE analysis was conducted for each vocabulary test. The analysis aimed to determine if the treatment groups demonstrated significantly greater learning gains on the vocabulary tests compared to the control group. Seven participants were excluded from the form recognition analysis for selecting three or more non-words in the post-test (Peters, 2019). Table 3 shows that both experimental groups demonstrated learning gains from pre-test to post-test, whereas the control group did not.

The GEE analysis for form recognition was performed on 470 observations, representing the target items that were unknown in the pre-test and were therefore available for learning in the post-test. The GEE analysis revealed that both intervention groups significantly outperformed the control group. The odds of recognizing a word's form were 3.95 times higher for the L2 subtitles group ($p = .003$) and 7.91 times higher for the L1 subtitles group ($p < .001$) compared to the control group.

Table 3
Form Recognition Scores

Condition	Pre-test score (Max = 32) Mean (SD)	Post-test score (Max = 32) Mean (SD)
Control (n = 7)	18.1 (3.08)	17.4 (5.47)
EngSub (n = 12)	21.5 (4.60)	24.6 (4.85)
SpaSub (n = 17)	18.4 (4.50)	20.1 (5.09)

Table 4 shows that while all groups demonstrated improvements, the largest improvements were observed in the L2 subtitles group.

Table 4
Meaning Recall Scores

Condition	Pre-test score (Max = 32) Mean (SD)	Post-test score (Max = 32) Mean (SD)
Control (n= 10)	9.80 (3.65)	10.7 (3.97)
EngSub (n= 15)	12.2 (4.87)	15.8 (5.72)
SpaSub (n= 18)	11.8 (4.89)	13.9 (5.06)

Table 5
Meaning Recognition Scores

Condition	Pre-test score (Max = 32) Mean (SD)	Post-test score (Max = 32) Mean (SD)
Control (n = 10)	19.1 (5.09)	19.5 (4.58)
EngSub (n = 15)	21.5 (4.14)	22.8 (4.43)
SpaSub (n = 18)	19.4 (5.59)	21.2 (5.81)

A second GEE analysis was conducted for the meaning recall test, performed on a total of 881 observations. The L2 subtitles group demonstrated significantly greater gains than the control group,

with the odds of recalling a word's meaning being 2.45 times higher ($p < .001$). In contrast, the effect for the L1 subtitles group was not statistically significant ($p = .068$).Table 5 indicates that all three groups improved from pre-test to post-test. A third GEE analysis was conducted on a total of 512 observations. No significant differences were found in learning gains between either of the subtitling conditions and the control group ($p = .624$).

6.3 The effect of visual support on incidental vocabulary learning

To address the first research question, the effect of visual support was examined across the three vocabulary tests. The results varied depending on the type of learning measured. For form recognition, the GEE analysis revealed a significant main effect for visual support (Wald $\chi2(1) = 6.28$, $p = .012$). However, the effect was contrary to expectations, as the odds of learning a word without visual support were 2.46 times higher than for a word with visual support (Table 6). In contrast, the GEE model confirmed that visual support was not a significant predictor for meaning recall ($p = .771$) (Table 7) or meaning recognition ($p = .800$) (Table 8).

6.4 The interaction effect of visual support and subtitling conditions

To address the third research question, the interaction between visual support and the subtitling condition was analyzed. The GEE analyses showed that this interaction was not significant for any of the vocabulary measures: form recognition (Wald $\chi2(1) = 1.13$, $p = .287$), meaning recall ($p = .975$), or meaning recognition ($p = .721$). This indicates that the effect of visual support (or lack thereof) did not differ significantly between the L1 and L2 subtitling groups.

6.5 Other significant predictors

While not the primary focus of the research questions, the analyses revealed other significant predictors of incidental vocabulary learning. To begin with, prior vocabulary knowledge was a significant predictor for both form recognition (Wald $\chi2(1) = 5.10$, $p = .024$), as it can be seen in Table 6, and meaning recall (Wald $\chi2(1) = 14.45$, $p < .001$) (Table 7). For every one-point increase in a participant's vocabulary score, the odds of learning a new word increased by approximately 4%. Cognateness was the strongest predictor for learning meaning. The odds of learning a cognate were over five times greater than for a non-cognate in the meaning recall test (Wald $\chi2(1) = 49.59$, $p < .001$) (Table 7) and 4.33 times higher in the meaning recognition test (Wald $\chi2(1) = 16.94$, $p < .001$) (Table 8). Finally, a significant main effect for subtitling condition was found for meaning recall (Wald $\chi2(1) = 4.89$, $p = .027$) (Table 7), where the odds of learning a word were 1.55 times higher for the L2 subtitles group compared to the L1 subtitles group. The summary of the findings are presented in Table 9.

Table 6
Summary of GEE Analysis for Predictors of Form Recognition

Predictor	**B**	**SE**	**Wald χ^2**	**p**	**OR**	**95% CI for OR**
(Intercept)	-3.00	1.25	5.79	.016	0.05	[0.00, 0.57]
Visual Support						
Without Visual Support vs. With	0.90	0.36	6.28	.012	2.46	[1.22, 4.99]
Subtitling Condition						
L1 Subtitles vs. L2 Subtitles	-0.46	0.46	1.00	.319	0.63	[0.25, 1.56]
3k UVLT Score	0.04	0.02	5.10	.024	1.04	[1.01, 1.07]
Visual Support × Subtitling Condition	0.45	0.42	1.13	.287	1.57	[0.68, 3.60]

Note. The analysis was based on 358 observations from 29 participants. *B* = unstandardized coefficient; *SE* = standard error; *OR* = odds ratio; CI = confidence interval.

Table 7
Summary of GEE Analysis for Predictors of Meaning Recall

Predictor	**B**	**SE**	**Wald χ^2**	**p**	**OR**	**95% CI for OR**
(Intercept)	-3.60	0.91	15.82	< .001	0.03	[0.01, 0.16]
Cognateness						
Non-cognate vs. Cognate	-1.66	0.24	49.59	< .001	0.19	[0.12, 0.30]
Subtitling Condition						
L2 subtitles vs. L1 subtitles	0.44	0.20	4.89	.027	1.55	[1.05, 2.28]
3k UVLT Score	0.04	0.01	14.45	< .001	1.04	[1.02, 1.07]

Note. The analysis was based on 659 observations from 33 participants. *B* = unstandardized coefficient; *SE* = standard error; *OR* = odds ratio; CI = confidence interval.

Table 8
Summary of the Final GEE Model for Predictors of Meaning Recognition

Predictor	**B**	**SE**	**Wald χ^2**	**p**	**OR**	**95% CI for OR**
(Intercept)	0.03	0.33	0.01	.918	1.04	[0.54, 1.99]
Cognateness						
Non-cognate vs. Cognate	-1.47	0.36	16.94	< .001	0.23	[0.12, 0.46]

Note. The analysis was based on 383 observations from 33 participants. *B* = unstandardized coefficient; *SE* = standard error; *OR* = odds ratio; CI = confidence interval.

Table 9
Summary of Main Findings by Research Question

Research Question	**Vocabulary Measure**	**Summary of Significant Findings**
RQ1: Does subtitling (L1 vs. L2) affect learning compared to a control group?	**Form Recognition**	Both L1 Spanish subtitles and L2 English captions were significantly more effective than the control group.
	Meaning Recall	Only L2 English captions were significantly more effective than the control group.
	Meaning Recognition	Neither subtitling condition was significantly more effective than the control group.
RQ2: Does visual support affect learning?	**Form Recognition**	Words *without* visual support were significantly more likely to be learned.
	Meaning Recall & Recognition	Visual support was not a significant predictor.
RQ3: Is there an interaction between visual support and subtitling condition?	**All Tests**	No significant interaction effect was found.
Other Predictors	**Across Multiple Tests**	Prior vocabulary knowledge and cognateness were strong and significant positive predictors of learning.

7 Discussion

7.1 Effects of L1 and L2 subtitles on incidental vocabulary learning from audiovisual input

The first research question examined whether subtitled input led to learning gains beyond test-exposure effects. The findings reveal that the intervention's effectiveness depended on the level of vocabulary knowledge assessed. For form recognition, both L1 subtitles and L2 captions significantly outperformed the control group. This confirms that audiovisual input with on-screen text facilitates the initial learning of word forms, a finding consistent with previous research (e.g. Peters, 2019).

A different pattern emerged for meaning recall, where only the L2 captions group significantly outperformed the control group. This suggests that L2 captions prompted deeper cognitive processing. By presenting the spoken and written forms in the same language, L2 captions may have encouraged a more effortful mapping of form to meaning from the visual context. Conversely, L1 subtitles, by providing an immediate translation, might have allowed for a shallower processing strategy where learners relied on the L1 text for comprehension, thereby bypassing deeper engagement with the L2 phonological form (Fievez et al., 2020; Peters, 2019).

Finally, for meaning recognition, neither subtitling condition proved more effective than the control group. This null result is likely attributable to the test's design and a potential ceiling effect. The high pre-test scores across all groups left little room for statistically detectable improvement, and the multiple-choice format may have provided all groups with strong cues that masked subtle learning gains (Fievez et al., 2020). Taken together, these findings suggest that the benefits of on-screen text are most apparent when the learning measure is sufficiently demanding to reflect differences in processing depth.

7.2 The effect of visual support in subtitled audiovisual input on incidental vocabulary learning

A key and surprising finding was the general lack of a facilitative effect for visual support, which contrasts with the existing literature (Fievez et al., 2020; Peters, 2019; Pujadas & Muñoz, 2023). For meaning recall and recognition, visual support was not a significant predictor. More strikingly, for form recognition, words without visual support were significantly more likely to be learned. This unexpected pattern may be attributable to several factors.

First, the lower-proficiency learners, evidenced by their lack of mastery for the levels beyond the first level in the UVLT, may have relied heavily on the on-screen text, ignoring the visual channel (Li, 2025). Second, the study's methodological design, which controlled for frequency by using only single-occurrence items, provides a direct test of imagery's effect. The null and negative findings here suggest that the benefits of imagery reported in prior research might have been amplified by the confounding effect of repeated verbal exposure for visually supported words. Finally, the powerful influence of other variables, particularly cognateness, may have statistically overshadowed any effects of imagery.

7.3 Other facilitative factors for word learning from subtitled audiovisual input

While the role of visual support was contrary to expectations, other findings aligned with and extended previous research. For the meaning recall, the finding that L2 subtitles were significantly more effective than L1 subtitles also aligns with previous research (e.g. Fievez et al., 2020; Peters, 2019). Captions may facilitate deeper learning in several ways. They provide learners with an exact written transcript of the audio, which aids in speech segmentation, helping learners parse the speech stream into individual words and making unfamiliar items more noticeable (Peters, 2019). This dual exposure to both the spoken and written L2 form creates a stronger memory trace (Mayer, 2021; Sadoski & Paivio, 2013).

Furthermore, prior vocabulary knowledge and cognateness emerged as powerful predictors of learning. The odds of recalling a cognate's meaning were over five times higher than for a non-cognate (Fievez et al., 2020; Peters, 2019; Peters & Webb, 2018), and the positive effect of a larger vocabulary size was comparable in magnitude to previous findings (Peters, 2019; Peters & Webb, 2018).

7.4 Pedagogical implications

The findings of this study, when situated within the broader context of research on vocabulary learning through audiovisual input, offer several practical implications for language teachers, materials designers, and learners. This study confirms that incidental vocabulary learning occurs through viewing captioned media, reinforcing the value of extensive viewing as a supplement to formal instruction (Webb, 2020). The effect of prior vocabulary knowledge underscores the principle that extensive viewing is most effective when learners already possess enough vocabulary. Webb (2020) argues that the accumulation of vocabulary through incidental vocabulary learning is a slow process. Therefore, teachers may catalyze vocabulary development through intentional learning activities to accommodate incidental learning through extensive viewing. For example, the facilitative effect of cognateness across multiple tests provides a clear directive for teachers. Explicitly teaching learners to recognize cognates between their L1 and the target language is a high-impact strategy that can significantly accelerate their vocabulary development as well as incidental vocabulary learning through extensive viewing by giving them immediate access to a large number of words (Peters & Webb, 2018). In conclusion, while incidental learning from audiovisual input is a valuable and motivating pathway for vocabulary growth, its effectiveness can be significantly enhanced through carefully selected pedagogical choices regarding on-screen text, a critical evaluation of visual support, and a focus on building learners' base knowledge and strategic awareness.

8 Conclusion

This study primarily investigated the role of visual imagery and on-screen text in incidental vocabulary learning in addition to other potential moderating factors. A key finding was that, contrary to the literature, visual support did not provide a learning benefit and was even found to be detrimental to form recognition. This suggests that for single-exposure items, the relationship between imagery and learning is complex; a single co-occurrence may be insufficient for learning to occur, and its effectiveness may depend on more factors like longer duration on screen. While the role of imagery was complex, L2 captions proved superior for meaning recall by fostering deeper cognitive processing than L1 subtitles. The study, though revealing, has some limitations. To begin with, the sample size is small to draw conclusions with great confidence. However, this limitation could be argued to have been minimized by having relatively larger number of target items (Brysbaert, 2025). Another limitation is due to the rather low proficiency of the participants. Some of the participants did not master even the first 1,000 word family level and majority did not master the second 1,000 word family level. However, such picture is a reflection of the English proficiency level in Spain where proficiency is not very high compared to other European countries, ranking 26 out of 35 (EF English Proficiency Index, 2024). Therefore, a future study recruiting higher proficiency learners might yield a more positive picture with respect to the effects of imagery.

Acknowledgements

This research was supported by the AGAUR-FI predoctoral program (2023 FI-1 00662) Joan Oró, funded by the Secretariat of Universities and Research of the Department of Research and Universities of the Government of Catalonia, and by the European Social Fund Plus.

References

Ahrabi Fakhr, M., Borzabadi Farahani, D. & Khomeijani Farahani, A.A. (2021). Incidental vocabulary learning and retention from audiovisual input and factors affecting them. *English Teaching & Learning (45),* 167–188. https://doi.org/10.1007/s42321-020-00066-y

Brysbaert, M. (2025). Applying mixed-effects models in research on second language acquisition: A tutorial for beginners. *Languages, 10*(2), 20. https://doi.org/10.3390/languages10020020

EF EPI Index. (2024). EF English proficiency index: A ranking of 116 countries and regions by English skills. https://www.ef.com/assetscdn/WIBIwq6RdJvcD9bc8RMd/cefcom-epi-site/reports/2024/ef-epi-2024-english.pdf

De Wilde, V., Brysbaert, M., & Eyckmans, J. (2020). Learning English through out-of-school exposure. Which levels of language proficiency are attained and which types of input are important? *Bilingualism: Language and Cognition*, *23*(1), 171–185. https://doi.org/10.1017/S1366728918001062

De Wilde, V., & Eyckmans, J. (2017). Game on! Young learners' incidental language learning of English prior to instruction. *Studies in Second Language Learning and Teaching, 7*(4), 673–694. https://doi.org/10.14746/ssllt.2017.7.4.6

Durbahn, M., Rodgers, M., Macis, M., & Peters, E. (2024). Lexical coverage in L1 and L2 viewing comprehension. *Studies in Second Language Acquisition*, *46*(4), 1045–1068. https://doi.org/10.1017/S0272263124000391

European Commission, Directorate-General for Communication, & Kantar. (2022). *Media use in the European Union: Report*. European Commission. https://data.europa.eu/doi/10.2775/911712

Fievez, I., Perez, M. M., Cornillie, F., & Desmet, P. (2020). Vocabulary learning through viewing captioned or subtitled videos and the role of learner-and word-related factors. *CALICO Journal, 37*(3), 233–253. https://doi.org/10.1558/cj.39370

Gyllstad, H., Sundqvist, P., Peters, E., Rindal, U., Skar, G., & Ulfat, N. (2025). Young learners' receptive L2 English vocabulary knowledge in relation to extramural English exposure at the onset of formal instruction in Norway. *International Review of Applied Linguistics in Language Teaching*. Advance online publication. https://doi.org/10.1515/iral-2024-0289

Howard, C. (Producer). (2023, March 19). Woodland (Season 1, Episode 2) [TV series episode]. In A. Fothergill (Executive Producer), *Wild Isles*. Silverback Films. Amazon Prime Video.

Koolstra, C. M., & Beentjes, J. W. (1999). Children's vocabulary acquisition in a foreign language through watching subtitled television programs at home. *Educational Technology Research and Development, 47*(1), 51–60. https://doi.org/10.1007/BF02299476

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

Leona, N. L., van Koert, M. J., van der Molen, M. W., Rispens, J. E., Tijms, J., & Snellings, P. (2021). Explaining individual differences in young English language learners' vocabulary knowledge: The role of extramural English exposure and motivation. *System, 96*, Article 102402. https://doi.org/10.1016/j.system.2020.102402

Li, Y. (2025). Listen or read? The impact of proficiency and visual complexity on learners' reliance on captions. *Behavioral Sciences, 15*(4), 542. https://doi.org/10.3390/bs15040542

Mayer, R. E. (2021). Cognitive theory of multimedia learning. In R. E. Mayer & L. Fiorella (Eds.), *The Cambridge handbook of multimedia learning* (pp. 57–72). Cambridge University Press. https://doi.org/10.1017/9781108894333.008

Nation, P., & Webb, S. A. (2011). *Researching and analyzing vocabulary.* Heinle.

Peters, E. (2018). The effect of out-of-class exposure to English language media on learners' vocabulary knowledge. *ITL-International Journal of Applied Linguistics, 169*(1), 142–167. https://doi.org/10.1075/itl.00010.pet

Peters, E. (2019). The effect of imagery and on-screen text on foreign language vocabulary learning from audiovisual input. *TESOL Quarterly, 53*(4), 1008–1032. https://doi.org/10.1002/tesq.531

Peters, E., & Webb, S. (2018). Incidental vocabulary acquisition through viewing L2 television and factors that affect learning. *Studies in second language acquisition*, *40*(3), 551-577. https://doi.org/10.1017/S0272263117000407

Pujadas, G., & Muñoz, C. (2023). Measuring the visual in audio-visual input: The effects of imagery in vocabulary learning through TV viewing. *ITL–International Journal of Applied Linguistics, 174*(2), 263–290. https://doi.org/10.1075/itl.22019.puj

Reynolds, B. L., Cui, Y., Kao, C.-W., & Thomas, N. (2022). Vocabulary acquisition through viewing captioned and subtitled video: A scoping review and meta-analysis. *Systems, 10*(5), Article 133. https://doi.org/10.3390/systems10050133

Rodgers, M. P. H. (2018). The images in television programs and the potential for learning unknown words: The relationship between on-screen imagery and vocabulary. *ITL - International Journal of Applied Linguistics, 169*(1), 191–211. https://doi.org/10.1075/itl.00012.rod

Sadoski, M., & Paivio, A. (2013). *Imagery and text: A dual coding theory of reading and writing* (2nd ed.). Routledge. https://doi.org/10.4324/9780203801932

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

Teng, M. F. (2024). Working memory and prior vocabulary knowledge in incidental vocabulary learning from listening, reading, reading-while-listening, and viewing captioned videos. *System, 124*, Article 103381. https://doi.org/10.1016/j.system.2024.103381

Teng, M. F. (2025a). Incidental vocabulary learning from listening, reading, and viewing captioned videos: Frequency and prior vocabulary knowledge. *Applied Linguistics Review, 16*(1), 477–507. https://doi.org/10.1515/applirev-2023-0106

Teng, M. F. (2025b). Modality of input and factors affecting incidental vocabulary learning: Reading, listening, and viewing with captions. *Applied Linguistics Review, 16*(4), 1607–1635. https://doi.org/10.1515/applirev-2024-0021

Uztosun, M. S., & Sundqvist, P. (2025). Extramural English activities and their relationship with L2 English proficiency at a Turkish university context. *The Literacy Trek, 11*(1), 56–78. https://doi.org/10.47216/literacytrek.1691753

Webb, S. (2020). Incidental vocabulary learning. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 225–239). Routledge. https://doi.org/10.4324/9780429291586

Webb, S., Sasao, Y., & Ballance, O. (2017). The updated Vocabulary Levels Test. *ITL-International Journal of Applied Linguistics, 168*(1), 33–69. https://doi.org/10.1075/itl.168.1.02web

Webb, S., Uchihara, T., & Yanagisawa, A. (2023). How effective is second language incidental vocabulary learning? A meta-analysis. *Language Teaching, 56*(2), 161–180. https://doi.org/10.1017/S0261444822000507

Kadir Kaderoğlu is a PhD candidate in the Cognitive Science and Language program at the University of Barcelona. His research interests are informal language learning, vocabulary learning from multimodal

input, and corpus linguistics. He has previously published in *CALL-EJ* and *ELIA: Estudios de Lingüística Inglesa Aplicada* (*Studies in Applied English Linguistics*). ORCID: 0000-0001-6974-054X