

Effects of Glossed Subtitles on L2 Vocabulary Acquisition

Yidan Wang, Bing Yan, Minh Tien Nguyen

Jiangsu Second Normal University, Nanjing 211200, Jiangsu, China

Copyright: © 2026 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

Abstract: Adopting a mixed experimental approach, this study compares the effects of L1-glossed keyword subtitles, L2-glossed keyword subtitles and non-glossed keyword subtitles on incidental L2 vocabulary acquisition. Thirty CET-4 undergraduates were randomly assigned to three groups. Data were collected through immediate vocabulary tests, three-week delayed retention tests, eye-tracking experiments and Likert questionnaires. One-way ANOVA reveals a clear hierarchical facilitating effect: L1-glossed subtitles rank first, followed by L2-glossed subtitles, and non-glossed subtitles perform worst, with significant gaps found in orthographic recognition, meaning recognition and meaning recall. Eye-tracking results indicate L1 glosses greatly reduce fixation duration and counts on target words and subtitle areas to optimize visual attention allocation. Correlation analysis proves appropriate focus on target words improves vocabulary gains, while excessive scanning of subtitle regions hinders learning. Delayed tests show L1-glossed subtitles produce the lowest forgetting rate and best long-term retention. Questionnaire data also demonstrate learners hold the highest recognition, usability evaluation and willingness to use L1-glossed subtitles. This study verifies that L1-glossed keyword subtitles reduce cognitive load, deepen form-meaning connections, and promote both short-term acquisition and long-term retention of L2 vocabulary.

Keywords: Glossed subtitles; Keyword subtitles; Incidental vocabulary acquisition; Eye-tracking

Online publication: April 26, 2026

1. Introduction

Videos are widely adopted to facilitate vocabulary learning due to rich contextual input, abundant lexis, high repetition of low-frequency words and visual scaffolding ^[1]. The core of vocabulary acquisition lies in establishing form-meaning mappings ^[2]. Subtitled videos are regarded as immersive, enjoyable contexts for lexical development, and subtitles function as effective learning supports for vocabulary growth. Scholars argue that subtitles boost lexical learning for several reasons: they encourage intentional noticing of linguistic forms, especially novel or distinctive expressions; they help segment continuous auditory streams into meaningful phonological units during listening decoding ^[3], they strengthen contextual cues for inferring word meanings; they sustain attentional resources critical for linguistic decoding; and they assist learners in constructing initial form-meaning connections, bridging phonological and orthographical representations of words to facilitate dual-modal lexical accumulation ^[4].

Prior research comparing full L2 subtitles and L2 keyword subtitles found that prominent keyword subtitles only improved orthographic recognition without generating significant advantages in meaning acquisition. Against this gap,

the present study explores optimized glossed subtitle formats to promote learners' comprehension and retention of word meanings^[5].

2. Research questions and participants

This study aims to explore how three subtitle variants (L1-glossed keyword subtitles, L2-glossed keyword subtitles, non-glossed keyword subtitles) affect L2 incidental vocabulary learning, with four specific research questions proposed:

- (1) Are there significant differences in target word orthographic recognition among learners exposed to L1-glossed, L2-glossed and non-glossed keyword subtitles?
- (2) Do the three subtitle types exert distinct effects on learners' ability to recognize lexical meanings?
- (3) How do the three subtitle formats differ in facilitating long-term recall of word meanings?
- (4) Do gloss conditions modulate learners' visual attention to target words and subtitle regions? Is there a correlation between attentional patterns and incidental vocabulary gains?

Thirty undergraduate students from a public university participated in the experiment (13 males, 17 females). All participants were native Chinese speakers with an average age of 18.96, having studied English for over ten years and passed the CET-4 examination. All participants had normal or corrected vision. Subjects were randomly assigned into three equal groups: L1-glossed group, L2-glossed group, and non-glossed group. Pre-tests measuring English proficiency and prior lexical knowledge were administered before the experiment. Independent samples t-tests confirmed no significant intergroup differences ($p > 0.05$), ensuring group equivalence.

3. Experimental materials

A 25-minute animated historical clip was selected as the stimulus material, depicting Spanish explorers' pursuit of the "golden dream" and their destruction of ancient American civilizations between the 15th and 17th centuries. Three English professors independently screened target keywords based on thematic relevance, lexical difficulty and frequency. Selected items mainly included nouns, verbs, adverbs and adjectives, accounting for roughly one third of all lexis in the script. To guarantee all target items were unfamiliar to participants, real words were replaced with pseudowords from the ARC pseudoword database. To control repetition as a confounding variable, each target pseudoword appeared exactly once throughout the video.

4. Research instruments

Four instruments were adopted: immediate vocabulary test, delayed retention test, eye-tracking equipment and closed-ended questionnaire.

The immediate test contained three subsections measuring orthographic recognition, meaning recognition and meaning recall, with 10 items per section and a full score of 30 (1 point per correct response). Orthographic recognition: Participants identified 10 target pseudowords among 20 mixed distractors. Meaning recognition: Each target word was paired with four Chinese semantic options (one correct, three semantically related distractors). Meaning recall: Participants wrote down the Chinese definition of each target word. Cronbach's $\alpha = 0.79$, indicating satisfactory reliability.

Administered two weeks after the immediate test, the delayed test reused identical target pseudowords with randomized item order, following the same scoring rubric (full score = 30). Cronbach's $\alpha = 0.76$, demonstrating reliable measurement of long-term lexical retention.

A 5-point Likert-scale questionnaire investigated learners' attitudes toward glossed subtitles, perceived utility in vocabulary learning and willingness to use such formats in future study. Cronbach's $\alpha = 0.82$, reflecting stable

measurement of subjective perceptions.

The SR Research EyeLink 1000 eye tracker with a sampling rate of 1000 Hz was used alongside a 21-inch 1024 × 768 display. Binocular tracking was enabled, while only left-eye data were recorded. Following Liversedge et al., fixations shorter than 80 ms or longer than 1200 ms were discarded as invalid outliers. Target words and subtitle areas were defined as Areas of Interest (AOIs). Four eye-movement metrics were analyzed: total fixation duration, first-pass fixation time, dwell time and fixation count. All data were processed via SPSS 22.0 with $\alpha = 0.05$ as the significance threshold.

5. Experimental design and procedure

A mixed-methods design was implemented, integrating quantitative lexical and eye-tracking data with qualitative questionnaire responses. All sessions took place in a quiet, well-lit language laboratory with individual workstations. One week prior to the experiment, all participants completed proficiency and lexical pre-tests to verify group homogeneity.

Participants sat 65 cm from the display and completed a 9-point grid eye-tracking calibration. Each participant watched the assigned video twice without pausing or rewinding; note-taking was prohibited to ensure incidental learning. Eye movements were recorded throughout playback. Immediately after viewing, participants completed the immediate vocabulary test and questionnaire (total session length ≈ 1 hour). Three weeks after the main experiment, all participants returned for the delayed retention test (≈ 50 minutes).

6. Results and discussion

One-way ANOVA was used to compare group differences across orthographic recognition, meaning recognition and meaning recall. Post-hoc Tukey tests were conducted when the main effect reached significance ($p < 0.05$). Parametric tests were applied after confirming normality and homogeneity of variance (Table 1).

Table 1. Descriptive statistics & one-way ANOVA for orthographic recognition

Group	N	M	SD	F	<i>p</i>	η^2
L1-glossed keyword subtitle group	10	16.62	3.28	7.892	0.002	0.386
L2-glossed keyword subtitle group	10	16.43	2.18			
Non-glossed keyword subtitle group	10	15.76	2.95			

The one-way ANOVA yielded a significant main effect of subtitle type ($F(2,27) = 7.892$, $p = 0.002$, $\eta^2 = 0.386$, medium-to-large effect size). Tukey comparisons revealed that the L1-glossed group outperformed the non-glossed group significantly (mean difference = 0.86, $p = 0.011$). No significant gap existed between L1-glossed and L2-glossed groups (mean difference = 0.19, $p = 0.785$), whereas the L2-glossed group scored markedly higher than the non-glossed group (mean difference = 0.67, $p = 0.013$).

Glosses enhanced orthographic processing relative to no glosses. Keyword subtitles themselves highlight word forms by segmenting audio streams, and glosses do not distract orthographic processing; instead, the “target word + gloss” compound boosts visual salience. L1 glosses enable faster form-meaning mapping, reducing unfocused visual scanning and improving orthographic processing efficiency (Table 2).

Table 2. Descriptive statistics & one-way ANOVA for meaning recognition

Group	N	M	SD	F	<i>p</i>	η^2
L1-glossed keyword subtitle group	10	11.06	1.91	11.620	< 0.001	0.465

Group	N	M	SD	F	<i>p</i>	η^2
L2-glossed keyword subtitle group	10	7.28	3.35			
Non-glossed keyword subtitle group	10	1.17	0.95			

A significant main effect was identified ($F(2,27) = 11.620, p < 0.001, \eta^2 = 0.465$, large effect). Multiple comparisons showed L1-glossed > L2-glossed (mean difference = 3.78, $p < 0.001$), L2-glossed > non-glossed (mean difference = 6.11, $p < 0.001$), with clear hierarchical differentiation. The non-glossed group relied solely on contextual inference, often generating inaccurate form-meaning links and remaining at superficial orthographic processing. L2 glosses offered definitional cues to reduce inferential errors, while L1 glosses eliminated cross-linguistic conversion costs and enabled instant, stable meaning mapping, yielding superior meaning recognition performance (**Table 3**).

Table 3. Descriptive statistics & one-way ANOVA for meaning recall

Group	N	M	SD	F	<i>p</i>	η^2
L1-glossed keyword subtitle group	10	11.38	2.31	15.893	< 0.001	0.532
L2-glossed keyword subtitle group	10	7.19	2.93			
Non-glossed keyword subtitle group	10	5.00	2.51			

The main effect of subtitle type was highly significant ($F(2,27) = 15.893, p < 0.001, \eta^2 = 0.532$, large effect). Post-hoc tests indicated L1-glossed scores were significantly higher than L2-glossed (mean difference = 4.19, $p < 0.001$), and L2-glossed exceeded non-glossed (mean difference = 2.19, $p = 0.023$). Meaning recall requires robust memory traces for active retrieval. Non-glossed participants formed weak, unstable lexical associations prone to retrieval failure. L2 glosses demanded extra cross-conversion processing, limiting memory consolidation. L1 glosses built direct L2-form-to-L1-concept mappings via deep semantic elaboration, creating durable memory representations and facilitating effortless recall.

Across all three lexical dimensions, the performance hierarchy held constant: L1-glossed keyword subtitles > L2-glossed keyword subtitles > non-glossed keyword subtitles. Glosses function as powerful input enhancements by providing instant semantic feedback, shifting cognitive resources away from error-prone contextual inference toward consolidation of form-meaning bonds. L1 glosses further eliminate cross-linguistic conversion overhead, allowing deeper lexical elaboration within limited viewing time and yielding comprehensive lexical learning gains.

AOIs were defined as target words and subtitle regions. Non-parametric Mann-Whitney U tests were adopted where data violated normality, with $\alpha = 0.05$. Spearman correlation with Bonferroni correction explored links between eye metrics and lexical scores. Attentional efficiency principle: equivalent learning outcomes achieved with shorter fixations and fewer gazes indicate lower cognitive cost and superior processing efficiency (**Table 4**).

Table 4. Target word eye-movement metrics (mean $\pm$ SD) & Mann-Whitney U results

Eye Metric	L1-glossed Group	L2-glossed Group	Non-glossed Group	L1 vs L2 (Z/p)	L2 vs Non-glossed (Z/p)
Total fixation time (ms)	386.2 $\pm$ 52.7	512.5 $\pm$ 68.3	698.8 $\pm$ 89.5	4.975/< 0.001	4.218/< 0.001
First-pass fixation time (ms)	215.3 $\pm$ 31.2	289.6 $\pm$ 42.5	365.9 $\pm$ 58.7	4.927/< 0.001	3.896/< 0.001
Dwell time (ms)	352.6 $\pm$ 48.9	478.3 $\pm$ 62.1	654.2 $\pm$ 81.3	6.175/< 0.001	4.562/< 0.001
Fixation count	1.8 $\pm$ 0.4	2.5 $\pm$ 0.6	3.6 $\pm$ 0.8	4.255/< 0.001	3.987/< 0.001

All pairwise comparisons reached significance ($p < 0.001$). The non-glossed group invested massive attentional resources in repeated, prolonged fixations to infer meanings from context, reflecting low processing efficiency. L2 glosses

reduced inferential load, shortening fixation durations and counts. L1 glosses enabled instant semantic access, requiring only simple form-concept matching with minimal visual investment yet maximal lexical gains, embodying high attentional efficiency.

Table 5. Subtitle area eye-movement metrics (mean $\pm$ SD) & Mann-Whitney U results

Eye Metric	L1-glossed Group	L2-glossed Group	Non-glossed Group	L1 vs L2 (Z/p)	L2 vs Non-glossed (Z/p)
Total fixation time (ms)	1258.3 $\pm$ 156.7	1896.5 $\pm$ 218.9	2568.7 $\pm$ 289.5	5.012/< 0.001	4.325/< 0.001
Dwell time (ms)	1125.6 $\pm$ 138.9	1752.3 $\pm$ 198.7	2389.2 $\pm$ 265.3	4.897/< 0.001	4.186/< 0.001
Fixation count	5.2 $\pm$ 1.1	7.8 $\pm$ 1.5	10.5 $\pm$ 2.0	4.658/< 0.001	4.012/< 0.001

Significant intergroup differences existed for all metrics ($p < 0.001$). Non-glossed participants scanned entire subtitle blocks repeatedly to extract contextual clues, incurring heavy cognitive load. L2 glosses localized semantic information to target words and narrowed processing scope. L1 glosses removed the need for extensive contextual analysis, requiring only brief visual verification of form-meaning pairs. In-line glosses placed directly beneath target words minimized visual shifts without adding extra subtitle processing burden, guiding selective attention to key lexical items (Table 6).

Table 6. Spearman correlations between eye metrics and vocabulary scores (L1-glossed group)

Eye Metric	Orthographic Recognition (r/p)	Meaning Recognition (r/p)	Meaning Recall (r/p)
Target word total fixation time	0.42/0.012	0.45/0.008	0.47/0.006
Target word dwell time	0.31/0.071	0.40/0.015	0.43/0.010
Target word fixation count	0.42/0.012	0.44/0.009	0.46/0.007
Subtitle area total fixation time	-0.36/0.038	-0.41/0.013	-0.44/0.009
Subtitle area fixation count	-0.35/0.045	-0.40/0.015	-0.43/0.010

Target word fixation duration and count correlated positively with lexical performance, demonstrating that deliberate focused processing of target words facilitates form-meaning consolidation. Subtitle area fixations correlated negatively with test scores: fewer scans of general subtitle text indicated more efficient selective attention toward target lexis, consistent with the attentional efficiency hypothesis.

Table 7. Descriptive statistics & one-way ANOVA for delayed vocabulary test

Group	N	M	SD	F	p	η^2
L1-glossed keyword subtitle group	10	9.86	2.15			
L2-glossed keyword subtitle group	10	5.79	2.58	12.358	<0.001	0.482
Non-glossed keyword subtitle group	10	3.25	1.89			

The ANOVA main effect was highly significant ($F(2,27) = 12.358$, $p < 0.001$, $\eta^2 = 0.482$). Tukey tests revealed L1-glossed retention scores were significantly higher than L2-glossed (mean difference = 4.07, $p < 0.001$), and L2-glossed outperformed non-glossed (mean difference = 2.54, $p = 0.018$). All three groups experienced score decay after three weeks, yet the decay gradient differed sharply: L1-glossed (-13.3%) < L2-glossed (-19.5%) < non-glossed (-35.0%). L1 glosses integrate new L2 lexis into established L1 semantic networks, forming dual episodic-semantic memory traces resistant to forgetting. L2 glosses only create intra-L2 associations with weaker consolidation, while non-glossed shallow processing yields fragile short-term memory prone to rapid decay. The 5-point Likert questionnaire (Cronbach's $\alpha = 0.82$) measured

three subjective dimensions (Table 8).

Table 8. Subjective evaluation scores (mean $\pm$ SD)

Dimension	L1-glossed Group	L2-glossed Group	Non-glossed Group
Perceived learning effectiveness	4.62 $\pm$ 0.38	3.58 $\pm$ 0.52	3.05 $\pm$ 0.45
Perceived usability	4.75 $\pm$ 0.29	3.26 $\pm$ 0.48	3.89 $\pm$ 0.42
Future willingness to use	4.81 $\pm$ 0.25	3.19 $\pm$ 0.55	2.08 $\pm$ 0.71

Participants rated L1-glossed subtitles highest across all dimensions. They reported instant semantic access eliminated dictionary lookup and contextual guessing, creating smooth viewing experiences and stable memory traces. While non-glossed subtitles had simpler visual layouts, their lack of semantic cues limited learning gains. Learners expressed low willingness to continue using L2 or non-glossed formats due to unsatisfactory vocabulary outcomes. Subjective feedback aligned with quantitative lexical and eye-tracking data, confirming L1-glossed subtitles match learners' cognitive preferences with high practical acceptability.

7. Conclusion

This experiment systematically compares the effects of L1-glossed keyword subtitles, L2-glossed keyword subtitles and non-glossed keyword subtitles on incidental L2 vocabulary acquisition. A consistent performance hierarchy emerges across immediate orthographic recognition, meaning recognition, meaning recall and delayed retention: L1-glossed > L2-glossed > non-glossed subtitles. L1 glosses deliver both superior short-term lexical learning and long-term memory retention. Eye-tracking evidence uncovers the underlying cognitive mechanism: L1 glosses optimize visual attention allocation by diverting cognitive resources from exhaustive contextual scanning to targeted consolidation of form-meaning mappings, requiring far less fixation time and fewer gazes on both target words and subtitle regions. Correlation analysis validates that focused target-word attention predicts better lexical performance, while excessive subtitle-area scanning hinders acquisition. Questionnaire data further confirm learners' stronger recognition, usability perception and adoption intention toward L1-glossed subtitles. Overall, L1-glossed keyword subtitles reduce cognitive overhead, deepen lexical semantic elaboration, and represent an optimized multimodal scaffold for incidental L2 vocabulary learning.

Funding

Philosophy and Social Science Research Project for Universities by Jiangsu Provincial Department of Education, "A Study on Influencing Factors and Optimized Modes of Subtitle Use Among Chinese Foreign Language Learners" (Project No.: 2021SJA0599)

Disclosure statement

The authors declare no conflict of interest.

References

- [1] Rodgers MPH, 2018, Incidental Vocabulary Acquisition from Viewing English Language Television Programs. *Language Teaching Research*, 22(4): 469–483.
- [2] Schmitt N, 2010, *Researching Vocabulary: A Vocabulary Research Manual*. Palgrave Macmillan, Basingstoke.

- [3] Yeldham M, 2018, Viewing L2 Captioned Videos: What's in It for the Listener? *Computer Assisted Language Learning*, 31(4): 367–389.
- [4] Puimege E, Peters E, 2019, The Role of Contextual Factors in Incidental Vocabulary Acquisition from Video. *System*, 83: 102105.
- [5] Hyona J, 2010, The Use of Eye Movements in the Study of Multimedia Learning. *Learning and Instruction*, 20(2): 172–176.

Publisher's note

Whioce Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.