

INCIDENTAL VOCABULARY RECOGNITION EFFECTS OF SUBTITLED, CAPTIONED AND REVERSE SUBTITLED AUDIOVISUAL INPUT

Jana van der Kolk

Sara Feijoo

Universitat de Barcelona, España

Abstract: *This study compares effects of brief exposure to L1 subtitled, L2 captioned, and reverse subtitled audiovisual input on three aspects of vocabulary learning: meaning, form, and pronunciation of target language words. A within-subjects design was used, in which three video clips in the different viewing modes were shown to ten L1 Spanish participants who underwent a pre-test, post-test, and delayed post-test of English terms that each occurred in just one of the videos. The tests measured recognition of meaning, form, and pronunciation through translation into the L1, dictation, and reading the target words aloud with native speaker ratings, respectively. The findings show statistically significant gains in producing the accurate written form of vocabulary in reverse subtitled and L2 captioned video clips, and mixed results for the other variables, including statistically significant gains in pronunciation of vocabulary with L1 subtitles.*

Key words: *EFL, subtitles, captions, reverse subtitles, multimodal input.*

1. INTRODUCTION

Audiovisual material offers a potentially rich source of input for second language learning by exposing users to the authentic spoken form of a target language, as well as its written form when L2 captions or reverse subtitles are applied. Multimodal input activates visual and auditory processing channels, potentially benefitting the development of listening and reading skills, and raising questions about how viewing modes affect a multitude of variables, including content comprehension and acquisition of specific linguistic features (Mayer, 2020). This study examines the incidental vocabulary recognition effects of three distinct viewing modes made possible by audiovisual material: L1 subtitles, L2 captions, and reverse subtitles. For the purpose of this report, *L1 subtitling* is defined as using native language subtitles with a target language soundtrack; *L2 captioning* involves both soundtrack and subtitles in the target language; and *reverse subtitling* refers to a native language soundtrack with target language subtitles.

2. LITERATURE REVIEW

2.1. Authentic multimodal second language input

Individuals' reported preference for either subtitled or dubbed viewing modes is somewhat specific to their country of origin and the viewing conditions to which they are accustomed as a consequence of policies affecting communications and education (Koolstra et al., 2002). Key factors connected with viewing preferences are higher proficiency in the foreign language in question as well as high education levels, both of which correlate positively with reliance on subtitles over dubbing (Araujo & Costa, 2013; Kuppens, 2010).

Numerous studies have explored the language acquisition effects of audiovisual material with L1 subtitles, L2 captions, reverse subtitles, or without any of these, resulting in a vast corpus of research on the topic, much of which reveals learning benefits of multimodal input (Caruana, 2021; Montero Perez et al., 2015; Winke et al., 2010). Innovation and expansion regarding authentic audiovisual materials is promising for language users and educators alike, who may be able to adjust the speed as well as the comprehensive or summary nature of the subtitles. The technology offers options, agency, individual purpose, and learning strategy development, all of which can contribute favorably to language acquisition across a variety of contexts (Pemberton et al., 2005). Among adults in particular, foreign language learning is a popular leisure activity in some cultural contexts, and

To cite this article: Van der Kolk, J., & Feijoo, S. (2024). "Incidental vocabulary recognition effects of subtitled, captioned and reverse subtitled audiovisual input". *Revista de Lingüística y Lenguas Aplicadas*, 19, 218-230. <https://doi.org/10.4995/ryla.2024.18056>

Correspondence author: janavanderkolk@gmail.com



audiovisual material can support the development of this lifelong habit (Vanderplank, 2016). Audiovisual input may be used both in instructed contexts and informally by learners for entertainment; therefore, research such as the present study might usefully explore the incidental effects of various viewing modes; that is, without the influence of formal instruction.

2.2. Comprehension with multimodal input

Comprehensibility of input, a foundational principle well established in theories of second language acquisition, may be enhanced through the representation of the meaning or form through multiple input modes. Several studies in this area focus on content comprehension and multimodal input (Gass et al., 2019; Jelani & Boers, 2018; Markham & Peter, 2003; Markham et al., 2001; Rodgers & Webb, 2017; Teng, 2019). Previous research has concluded that L1 subtitles enhance comprehension when compared to both the absence of L1 subtitles and presence of L2 captions (Liao et al., 2020; Markham et al., 2001; Markham & Peter, 2003; Matielo et al., 2018; Pujadas & Muñoz, 2020; Vulchanova et al., 2015).

Various studies have concluded that L2 captions are a benefit for content comprehension, due to the interconnectedness of cognitive functions involved in processing multimodal input (Bird & Williams, 2002; d'Ydewalle, 2002; Montero Perez et al., 2015). Researchers have also found that L2 captions do not necessarily interfere with auditory processing (Winke et al., 2010). Viewers themselves have reported a perceived advantage from the addition of a textual input mode, particularly in instances where speech is too complex to grasp from auditory input alone (Pattemore et al., 2020; Sydorenko, 2010; Vanderplank, 2019).

Furthermore, other studies suggest that viewers have a general inclination to read any subtitles or captions regardless of whether it is necessary for comprehension or indeed familiarity with the language displayed; one's attention is inevitably drawn to text that appears on-screen (Bisson et al., 2015; Muñoz, 2017). This finding underscores the potential learning benefits of multimodal input, including in conditions such as reverse subtitling, in which the target language is made available only in written form.

2.3. Vocabulary acquisition with multimodal input

A wide variety of experimental studies have focused on both receptive and productive vocabulary gains (Mitterer & McQueen, 2009; Pujadas & Muñoz, 2019; Wang, 2019). Benefiting from authentic audiovisual input implies a threshold of vocabulary knowledge, and a helpful contribution to this area of research has been an effort to quantify the lexical variety presented by television shows and films (Webb & Rodgers, 2009a; 2009b). This analysis found that the variety of vocabulary differed considerably across genres and even between episodes; nevertheless, low-frequency vocabulary is rarely encountered in certain types of entertainment programming, especially compared with news reporting, making the former more accessible for non-expert language learners. This study recognized that vocabulary knowledge is the primary hurdle for viewers' comprehension and a prerequisite for coping with fast speech rates and deciphering the meaning of unfamiliar lexis from context. Positive correlations have been found between prior vocabulary knowledge and word recognition gains from subtitled multimedia (Fievez et al., 2020; Peters et al., 2016). Assuming a sufficient knowledge base, it is reasonable to predict that informal use of this resource would contribute to ongoing vocabulary learning (Webb & Rodgers, 2009a: 425).

Some research involving multimodal input has centered on vocabulary learning (Aydin Yildiz, 2017; Fazilatfar, Ghorbani & Samavarchi, 2011; Mardani & Najmabadi, 2016; Nasab & Motlagh, 2017; Sadiku, 2018; Yang, 2018) and grammar rule acquisition (Pattemore & Muñoz, 2020). The presence of L2 captions has been found to improve listening skills and word recognition (Hsieh, 2020). While meta-analyses encouragingly conclude that L2 captions are beneficial to listening comprehension and vocabulary acquisition (Caruana, 2021; Jaber & Dollar, 2019; Montero Perez et al., 2013), a common limitation of studies of multimodality is their focus on comparing the presence and absence of L1 subtitles or L2 captions, rather than addressing other multimodal input options such as reverse subtitling.

Certain studies involve learning contexts that could be described as informal and which result in incidental effects: recruiting uninstructed participants, providing no stated goals during interventions, allowing repeated viewing, and surveying participants' at-home use of subtitles (Garnier, 2014; Jelani & Boers, 2018; Vanderplank, 2019). Studies generally focus on features that occur repeatedly in the video with high degrees of visible and audible salience (Matielo et al., 2015: 162) and have found that repetition and visual representation correlate positively with vocabulary recognition gains (Fievez et al., 2020). Incidental vocabulary gains have been found in terms of recognition and recall of target language terms (Puimège & Peters, 2019). The present study aims to expand on these findings by including pronunciation as a dependent variable, measuring longitudinal retention, and examining the possibilities of reverse subtitling.

Valuable studies have found L1 subtitles to be a detriment to target language acquisition, and L2 captions to be a benefit, based on findings that L2 captions improved comprehension of regional accents, and also improved production of L2 speech (Mitterer & McQueen, 2009; Wisniewska & Mora, 2020). These researchers conclude that L2 captions lexically guide learners to retune their perceptual categories since captions are presented alongside phonological input. By contrast, this benefit is not available from L1 subtitles, which in fact appeared to hinder the adaptation of perceptual categories and assimilation of unfamiliar language features.

2.4. Reverse subtitling possibilities

Reverse subtitling has been omitted from many previous experimental designs focusing on multimodality, with a few notable exceptions (Danan, 1992; Garnier, 2014; Gorjian, 2014; Fazilatfar et al., 2011; Mardani & Najmabadi, 2016). Several existing studies of reverse subtitling have found acquisition gains, including in written production of target language vocabulary by beginner learners (Danan, 1992) and recognition of new vocabulary by children and adults of various proficiency levels (Gorjian, 2014).

Danan (1992) compared all three multimodal viewing conditions examined in the present study. Participants were tested on their production of the written form of vocabulary items from the input; however, the potential for pronunciation gains from the target-language soundtrack in the L1 subtitled and L2 captioned modes was not measured. Furthermore, participants were provided with summaries of the videos before viewing and given printed scripts of the videos that omitted the target terms. Therefore, the acquisition gains found in this study should be described as instructed rather than incidental.

A larger study of all three viewing modes using a similar within-subjects design also found gains from reverse subtitled viewing (Gorjian, 2014). However, this study measured gains only in terms of target terms' meanings. Furthermore, this study does not reflect incidental or informal acquisition, since it entailed multiple treatment sessions that included discussion and quizzing of video content as well as elicitation of target vocabulary outside of testing times. The study also lacked a measure of longitudinal retention.

Two large studies of all three viewing modes using a between-subjects design found the greatest vocabulary recognition gains from reverse subtitled viewing (Fazilatfar et al., 2011; Mardani & Najmabadi, 2016). These results, while encouraging, are based on tests of meaning recall, such as translation, that leave open questions about other potential benefits of reverse subtitling, such as whether it could improve learners' production of the written form of novel vocabulary. Another gap in the research is the frequent absence of measures of long-term retention of gains. Garnier (2014) found both vocabulary gains and long-term retention from reverse subtitled film viewing; however, as a case study of a single participant, these findings cannot be generalized.

A few broad generalizations emerge from the substantial existing research on multimodal input: the addition of input modes can enhance comprehensibility where the level of complexity is appropriate for the viewer; L1 subtitles have demonstrable benefits for content comprehension, while reverse subtitles and L2 captions may be useful for incidental lexical acquisition.

In addition to incorporating longitudinal measures, the present study operationalizes vocabulary recognition in terms of elicitation of written form and translation of target language items into participants' native language, which comprise a more rigorous and reliable measure of potential gains in vocabulary knowledge. The study also examines the possibility that target language auditory input could result in pronunciation gains by including a pronunciation test. To reflect language users' independent, informal practices, the experiment implements an uninstructed learning context involving authentic second language input using all three multimodal viewing options. Potential vocabulary recognition gains are further corroborated by measuring longitudinal retention. Content comprehension as well as participants' informal viewing habits and preferences are taken into consideration as relevant factors in understanding possible variability in their performance.

3. OBJECTIVES

This experiment measures gains in both receptive and productive vocabulary knowledge – meaning, form, and pronunciation – as well as the effects of different viewing modes on content comprehension. Participants were exposed to a type of input that they might consume in their leisure time, with just one viewing and no stated learning goals; the latter criteria constitute incidental rather than instructed learning environment. Specifically, the present study aimed to answer the following research questions:

- *RQ1*: What are the gains in recognition of meaning, form, and pronunciation of new vocabulary resulting from viewing L1 subtitled, L2 captioned, and reverse subtitled video?

- RQ2: Do L1 subtitled, L2 captioned, and reverse subtitled viewing correspond to different degrees of content comprehension on the part of viewers?
- RQ3: What relationships exist between viewers' levels of formal and informal target language exposure and their vocabulary recognition gains from watching L1 subtitled, L2 captioned, and reverse subtitled video?

Previous research of this kind invites the prediction that content comprehension is best facilitated by L1 subtitle use. Given existing findings showing the importance of corroborating meaning through multiple input channels, and the full availability of meaning in the L1 soundtrack of reverse subtitled video, it can be predicted that the reverse subtitled viewing condition would also produce high content comprehension. Conversely, among the three viewing conditions, L2 captioning is the least likely to favor comprehension.

Eye-tracking evidence that viewers inevitably pay attention to on-screen text suggests that reverse subtitling might contribute to recognition of the written form of target language items. On the other hand, pronunciation would be unlikely to improve in this condition given the absence of L2 auditory input, in contrast with the L1 subtitled and L2 captioned viewing conditions, which might produce pronunciation gains.

4. METHODOLOGY

4.1. Participants

The participants in this study were ten teenagers enrolled in an afterschool English language course at a private academy in Barcelona, Spain. They were all native Spanish/Catalan bilinguals, and English was their first foreign language. The group ranged in ages from 14 to 19 ($M = 15.5$, $SD = 1.78$). English proficiency tests upon entrance at the language school showed that their language proficiency levels ranged from high A2 to low B1 according to the Common European Framework of Reference for Languages.

4.2. Design

This study entailed a pre- and post-test design, with a delayed post-test. Knowledge of meaning, form, and pronunciation of a list of target vocabulary items were measured identically at each testing time. The pretest was conducted one week prior to the one-time viewing of three short videos, the post-test took place immediately after viewing the videos, and the delayed post-test occurred one month thereafter.

The authentic audiovisual material used in this experiment consisted of three ten-minute segments from different episodes of the American television series *Friends*. This series was selected for the accessibility of the situational comedy genre, in which episodes revolve around humor and simple subplots that viewers can follow without being familiar with the series. In terms of lexical coverage, in each of the three episodes used, the percentage of words belonging to the first 1000 most frequently used English words (K1) ranges from 70.6-76.4%; and the percentage of words belonging to the second 1000 most frequent words (K2) ranges from 10-14.4%. Lexical bands above K6 each contain fewer than 1% of the words encountered in any of the episodes (Compleat Web VP, 2021).

The selected clips consisted of the first ten minutes of each of three episodes, the span in which key actions and ideas are framed and presented. The videos were streamed via Netflix, which enables English captioning as well as dubbing and subtitling in European Spanish. When asked, none of the participants reported having previously watched the series.

Five or six vocabulary items were selected from each of the three video clips and used in tests of vocabulary recognition (appendix A). It was essential to establish that each item occurred in only one of the three clips and was translated or transmitted in both the corresponding text and soundtrack. Objects represented visually on-screen were not included. The seventeen target terms include verbs, adjectives, and nouns, ten of which occurred only once in the script, while six were repeated twice and one occurred three times. The final target word list also emerged from an initial consultation with the participants' English language teacher, who predicted the level of lexical complexity that would reduce the likelihood of participants' prior familiarity with the target terms. Ultimately, the pre-test phase measured participants' existing knowledge of the finalized list.

The video clips were shown consecutively and without introduction or instructions: the first clip was shown with subtitles, the second with captions, and the third with reverse subtitles. This experiment applied a within-subjects design in which all ten participants viewed all of the videos, with the aim of comparing differences in their vocabulary recognition outcomes across the three conditions.

4.3. Instruments

The first test delivered was a dictation of the target terms in which participants attempted to produce the correct written form of each term: each target term was repeated twice consecutively by the researcher. This test was conducted first in order to preclude participants seeing the written form of the target terms before this test, since they would see the terms in subsequent tests. Next, recall of the meaning of the target terms was tested through a translation task in which participants converted a written list of the vocabulary into Catalan or Spanish equivalents. The final vocabulary test measured pronunciation and entailed participants individually reading aloud the list of target terms into a voice recorder.

A comprehension test was conducted following the video viewing and consisted of fifteen true or false statements relating to major plot events, five from each video viewed (Appendix B). This test provided an indication of the extent to which each viewing condition facilitated comprehension.

Participants' linguistic background, target language media consumption, and subjective attitudes toward subtitling and dubbing were ascertained through short surveys following the pre- and post-tests (Appendix C). Participants indicated their overall target language exposure in formal and informal terms, and their relative preferences for using subtitled, captioned, and dubbed media, whether due to necessity or aesthetic preference. Responses were selected among multiple choices or on a six-point Likert scale.

4.4. Scoring

Responses on the dictation test of form were scored as valid or not based on whether they mirrored the completely correct written form of the term in question. Responses on the translation test of meaning were rated as correct or not depending on whether the translation provided fully conveyed an accurate meaning of the target term. Participants' recorded pronunciation of the target terms was rated by two native speakers of American English, both of whom are English language teachers with applied linguistics research experience. The raters used a six-point Likert-type scale ranging from *unintelligible* to *native-like*; averages of the two ratings served as measures of participants' pronunciation.

5. RESULTS

5.1. Descriptive data

The results of the vocabulary tests show improvement in average scores from the pre- to post-tests in all viewing conditions and across all acquisition variables (Table 1). For acquisition of form in particular, gains are comparatively greater in the reverse subtitled and captioned conditions. Gains on the delayed post-tests compared to pre-test scores show retention of vocabulary recognition in the longer term across all conditions.

Table 1. Comparison of mean scores (N = 10).

Viewing condition	Acquisition variable	Pre-test scores	Standard deviation	Post-test scores	Standard deviation	Delayed post-test scores	Standard deviation	Gains	Long-term gains
Subtitles	Meaning	0.70	0.82	0.90	0.74	0.88	0.74	0.20	0.18
	Form	1.40	1.08	2.10	1.20	2.00	1.20	0.70	0.60
	Pronunciation	2.23	0.42	2.42	0.48	2.39	0.49	0.19	0.16
Captions	Meaning	2.90	0.99	3.30	0.95	3.63	0.95	0.40	0.73
	Form	1.40	1.08	3.00	1.49	2.88	1.27	1.60	1.48
	Pronunciation	2.63	0.37	2.71	0.37	2.74	0.37	0.08	0.11
Reverse subtitles	Meaning	1.00	0.67	1.70	1.06	1.75	1.06	0.70	0.75
	Form	1.00	0.82	3.30	1.25	3.00	1.41	2.30	2.00
	Pronunciation	2.03	0.35	2.40	0.42	2.44	0.45	0.37	0.41

In the pre- and post-test comparison, the sample size referred to is the total of ten participants; however, delayed post-test scores from only eight participants are included, due to reduced attendance at the final testing session. A normal distribution of the results was not found. Therefore, nonparametric Wilcoxon signed ranks tests served to assess the significance of gains in participants' scores across the three testing times.

5.2. Research Question One: Form

The dictation post-test of the written form of the target terms reveals gains across all conditions, with significantly greater gains for terms encountered in reverse subtitled input ($Z = -2.724$, $p = 0.006$) and L2 captioned input ($Z = -2.831$, $p = 0.005$). In the L1 subtitled condition, by contrast, this difference was not found to be statistically significant ($Z = -1.511$, $p = 0.131$).

Scores for written form improved for virtually all participants in the reverse subtitled and L2 captioned conditions (nine and ten participants, respectively), and did not decrease for any. By way of comparison, the L1 subtitled viewing condition resulted in increasing form scores for four participants, with half of the group seeing no change in scores on the post-test in this condition.

5.3. Research Question One: Meaning

In the translation tests of the meaning of the target terms, the group's average scores increased from pre- to post-test in all viewing conditions (Table 1), but the terms encountered in reverse subtitled and L2 captioned input correspond to the greatest gains in correct responses on the post-test. The differences between the average scores on the pre- and post-tests of meaning do not attain statistical significance in any viewing condition ($Z = -1.823$, $p = 0.068$ in the reverse-subtitled condition; and $Z = -1.414$, $p = 0.157$ in both the L1 subtitled and L2 captioned conditions).

5.4. Research Question One: Pronunciation

The results of the pronunciation tests show larger gains for target terms encountered in L1 subtitled and reverse subtitled input. Average pronunciation scores of the group increased to some extent in all conditions but there were statistically significant differences in the L1 subtitled condition ($Z = -2.328$, $p = 0.02$) and reverse subtitled conditions ($Z = -2.257$, $p = 0.024$). Statistically significant differences were not found to result from L2 captioned viewing, on the other hand ($Z = -0.701$, $p = 0.483$).

Pronunciation scores increased for seven participants in both the L1 subtitled and reverse subtitled input conditions. In the L2 captioned viewing condition, on the other hand, pronunciation scores for the target terms encountered improved for only four participants.

5.5. Research Question One: Long-term retention of meaning, form and pronunciation

The results of the delayed post-test were also considered, in order to establish whether participants retained the vocabulary recognition gains from the multimodal input one month after viewing. Participants whose scores remained constant on both the post-test and the delayed post-test are assumed to have retained at least the same level of knowledge measured at the post-test. All recognition measures across all viewing conditions showed higher delayed post-test scores compared to the pre-tests, though some reduction in scores from the post- to delayed post-test scores can also be seen (Table 2).

Regarding form, average scores across all participants decreased slightly for terms encountered in each viewing condition from the post- to the delayed post-test. The delayed post-tests of form in the reverse subtitled condition resulted in a decreased score for three participants, compared with two participants whose scores dropped in the other two viewing conditions.

The results of the delayed post-test of meaning of the target terms show that scores decreased one month after the post-test for two participants in each viewing condition. A slight decrease is seen in the group's average scores in the L1 subtitled viewing condition over time ($M = 0.9$ and $M = 0.88$, respectively). By contrast, participants' average meaning scores for terms from the L2 captioned and reverse subtitled conditions improved slightly one month later.

Delayed post-test pronunciation scores fell for more participants in the reverse subtitled condition: four lower scores compared with two and three lower scores in the L2 captioned and L1 subtitled viewing conditions, respectively, when compared to the post-tests. Average pronunciation scores across all participants dropped slightly for terms encountered in the L1 subtitled viewing condition ($M = 2.42$ and $M = 2.39$). By contrast, average scores increased minutely in both the L2 captioned and reverse subtitled conditions.

Table 2. Wilcoxon Signed Ranks Test Statistics.

Mode	Variable	Pre- to post-test (N=10)		Pre- to delayed post-test (N=8)	
		Z-score	p value	Z-score	p value
Subtitles	Pronunciation	-2.328	.020	-2.124	.034
Captions	Form	-2.724	.006	-2.414	.016
Reverse subtitles	Form	-2.831	.005	-2.456	.014
	Pronunciation	-2.257	.024	-2.047	.041

5.6. Research Question Two: Comprehension

Results of the comprehension test (Appendix B) varied across the different subtitling conditions (Table 3). L1 subtitled viewing led to the fewest errors, with an average score of 4.7 out of five possible correct responses and the lowest variability as well (SD =0.483). L2 captioning produced the lowest average and most variable comprehension scores (M = 4.2, SD = .919), as well the lowest individual score of three out of five possible correct responses.

Table 3. Comprehension scores (N = 10).

Viewing condition	Mean score	Minimum score	Maximum score	Standard deviation
Subtitles	4.7	4	5	.48
Captions	4.2	3	5	.92
Reverse subtitles	4.4	4	5	.52

5.7. Research Question Three: Informal language exposure

The surveys we carried out provided insight into participants' attitudes toward subtitling and dubbing and their informal viewing habits. When asked what language, if any, they would prefer to see written on-screen, the most common preference was for Spanish text, with four out of ten participants choosing this option, and the remaining six responses evenly divided between English and no subtitles at all. Other viewing preferences were indicated on a six-point Likert-type scale with values ranging from 0 (*disagree or never*) to 5 (*agree or always*). These qualitative results show an overall preference for audiovisual material dubbed in the viewers' native language, Spanish or Catalan (M = 3.5). At the same time, participants showed comparable willingness to view material with either English or Spanish subtitles (M = 1.1 and M = 1.5) as well as considerable appreciation for viewing videos with their original soundtrack even if subtitles are required for comprehension (M = 2.9).

Table 4. Language Exposure & Viewing Preferences (N = 10).

Survey question/statement	Mean	Minimum	Maximum	Standard deviation
How many hours per week do you spend in English class and preparing for class?	5.85	3	10	2.16
How many hours per week do you use English-language media in your free time?	2.70	0	8	2.41
6-point Likert-type scale (0/never – 5/always):				
I use English subtitles when watching videos & films in my free time.	1.10	0	4	1.29
I use Spanish subtitles when watching videos & films in my free time.	1.50	0	4	1.58
I watch the Spanish dubbed version if it's available.	3.50	1	5	1.51

In an attempt to address the research question aimed at exploring the possible relationship between learners' formal and informal exposure to the target language and their vocabulary acquisition gains from viewing subtitled audiovisual input, Spearman's *r* correlations between these variables were largely found to be weak and non-significant. A moderate positive correlations was found between gains in pronunciation within the L2 captioned viewing condition and total language exposure, though without attaining statistical significance ($r = 0.536, p = 0.111, N = 10, R^2 = 0.29$). A Spearman's *r* correlation between total weekly hours of English exposure

and participants' degree of preference for Spanish dubbing as a viewing mode, indicated on a six-point Likert-type scale ranging from *never* to *always*, showed a statistically significant negative correlation with a large effect size and fairly wide confidence interval (95% CI: .40, .88; $r = -0.80$, $p = 0.005$, $N = 10$, $R^2 = 0.64$).

6. DISCUSSION

The results of this experiment reveal potential word recognition benefits of reverse subtitled and L2 captioned audiovisual media, particularly in terms of production of the written form of target language vocabulary and, to a lesser, statistically insignificant extent, recognition of meaning. Gains in producing written form of target terms from the reverse subtitled and L2 captioned videos are considerably greater than those resulting from the L1 subtitled video. As a within-subjects study involving an intact group of ten participants, the results, however compelling, are suggestive rather than generalizable.

Recognition of meaning acquired in the reverse subtitled condition could be due to its corroboration through the native language soundtrack available in this viewing mode. The gains found mirror those of much larger studies (Fazilatfar et al., 2011; Gorjian, 2014; Mardani & Najmabadi, 2016). However, these studies measured learning only in terms of recognition of target terms' meaning through their inclusion in multiple-choice questions or translation tasks. The present study's concept of recognition effects includes elicitation of written form and translation of target language items, comprising a multi-faceted measure of vocabulary learning, as well as taking the question of long-term retention into account with a delayed post-test. Nevertheless, it is worth noting the comparable results of the abovementioned studies since they involved Iranian learners of English and therefore a considerably larger typological and orthographical distance from the native language in question, Persian, while showing the same favorable outcomes for reverse subtitling.

The findings of the present study in terms of gains from reverse subtitled and L2 captioned viewing might be partly explained by the apparent tendency of viewers to pay attention to on-screen text. Gains on tests of meaning and particularly form within the reverse subtitled viewing condition echo the findings of previous research that also measured written production, albeit based on inauthentic audiovisual input aimed at novice learners (Danan, 1992). The latter study also included a measure of longitudinal retention; however, the time span between the post- and delayed post-test is not specified in that report.

The improvement of meaning and form scores in the L2 captioned viewing condition among this sample suggests that the comprehensibility of the input for the participants' target language proficiency level may have allowed them to infer the meaning of unfamiliar terms from the various input channels, without the benefit of any of the latter channels being available in their native language. On the other hand, the relatively lower comprehension score within this viewing mode, considered alongside the aforementioned vocabulary recognition gains, suggests that viewers' attention may have been strained overall, possibly contributing to a reduced general understanding of content.

Analysis of the pronunciation results might invite the belief that L1 subtitled and reverse subtitled viewing modes produce gains in native-like pronunciation of new vocabulary, since the differences in pre- and post-test scores in these conditions exceed differences found in the L2 captioned condition and attain statistical significance. These findings corroborate the conclusions of previous research that L2 captions provide orthographic cues indicating which target language items are being heard and therefore contribute to improved spoken production of new vocabulary to a greater extent than L1 subtitles (Mitterer & McQueen, 2009). However, target language auditory input is absent from the reverse subtitled viewing condition, which raises the question of whether the improvement in pronunciation found within this condition in the present study resulted from the audiovisual input itself or from other factors. In light of this, conclusions about pronunciation from this study should be interpreted with caution.

No significant correlation was found between participants' average target language exposure and their gains in any of the measures of vocabulary recognition across any of the viewing conditions, echoing the ambiguous findings of other related research (Sockett, 2014). This attempt to answer the research question regarding the effects of exposure on gains from multimodal viewing has therefore been inconclusive. Future research designs might attempt to disentangle these factors by incorporating measures of informal exposure such as immersion experience, specific aspects of proficiency like receptive language skills and aptitude, and comparison of media genres consumed informally with the effects of specific genres used in experimental interventions. In addition, measuring facets of learning beyond vocabulary recognition, such as familiarity with regional accents and pragmalinguistic features, might more clearly reveal a role of different types of exposure. Finally, the strong negative correlation between total current English exposure and reported preference for Spanish dubbing as a viewing mode is notable and suggests that media consumption choices by language users may relate to the status of their proficiency development, particularly receptive language skills, as well as affective factors (Pattimore et al., 2020; Suárez & Gesa, 2019).

The self-reported preferences of the participants in the present study suggest that a variety of viewing modes warrant study, with roughly equal numbers of participants reporting that they select English, Spanish, and no subtitles in their informal viewing habits. While it appears that participants were not entirely averse to seeing the target language on screen, these results do not demonstrate that reverse subtitling as opposed to L2 captioning is preferred by those who reported using English subtitles, since the survey questions do not specify the audio language used in this case (Appendix C).

The design of the pronunciation test poses a potential limitation to the present study. The three tests of vocabulary recognition were conducted with the small groups of participants in the same order at each testing time: first a dictation test, then a translation test, and finally the pronunciation test, the last of which was administered individually with each participant in a separate room, where they read the target terms aloud into a voice recorder. The dictation test, which was always conducted first, entailed the researcher repeating each target term aloud twice while the participants attempted to produce the correct written form. This was done in order to make sure that participants did not receive any additional exposure to the written form of the target words (i.e. by seeing them in the translation task or the pronunciation task) before the dictation test, which would have reduced the validity of the written form recognition measures.

However, as a result of the dictation task being carried out first, all participants heard the target terms spoken clearly before subsequently producing them during the pronunciation test. The study aimed to measure incidental word recognition gains from watching videos; however, the auditory input from the dictation test constitutes additional exposure apart from the videos themselves. This repetition of auditory cues may have impacted participants' subsequent improved spoken production of those terms, and could explain the improvement of participants' pronunciation of target terms encountered in the reverse subtitled video, where the native language soundtrack could not have been the source of input necessary to produce the pronunciation gains seen on the post-test. Given that the additional auditory cues from the dictation test potentially impacted all participants' performance equally, the pronunciation test results have been included in the final analysis. This testing issue could be avoided in similar research designs in the future by separating the pronunciation test from other vocabulary tests by several hours or even days.

Additional limitations of this study include the small sample size and a possible ceiling effect with the comprehension test. Future experiments of this kind might aim to involve more participants, as well as controlling for the influence of variables such as individuals' aptitude and proficiency in receptive and productive skills. Finally, the possible ceiling effect with the comprehension results complicates the effort to examine the effect of each viewing condition on this variable. This test might have yielded more insight if the questions posed were more challenging for the participants, leading to a wider range of scores within each viewing condition.

7. CONCLUSION

This study offers some suggestive findings in support of the potential of reverse subtitles and L2 captions as conduits of vocabulary recognition. In light of gains found in meaning and written form within the reverse subtitled mode, there is reason to question the prevailing notion that a target language soundtrack is a preferable option for developing foreign language competence. Reverse subtitling potentially offers a compromise between a general inclination toward choosing a native language soundtrack, obtained through dubbing if necessary, and a desire to access the potential language learning benefits of multimodal audiovisual materials. While pronunciation input is missing from the reverse subtitled condition, the reduced strain on comprehension enabled by the native language soundtrack may be valuable in itself, particularly for novice learners with less developed receptive skills in the second language, and warrants further study.

Research into the potential of subtitle and caption use for language learning is of particular interest to those seeking to expose themselves or other language users to authentic target language input in order to further their language proficiency development using materials created by and for native speakers. This study has attempted to contribute to the conversation by measuring vocabulary recognition and pronunciation in relation to three multimodal viewing options with distinct benefits.

ACKNOWLEDGEMENTS

This research was supported by grant PID2019-110594GB-I00 from the Spanish Ministry of Science and Innovation, and grant 2020FI_B2 00179 from the Catalan Agency for Management of University and Research Grants.

REFERENCES

- Araujo, L., & Costa, P. (2013). The European Survey on Language Competences: School-Internal and External Factors in Language Learning. *JRC Scientific and Technical Report* (EU 26078-2013). Luxembourg: Publications Office of the European Union.
- Aydin Yildiz, T. (2017). The Effect of Videos with Subtitles on Vocabulary Learning of EFL Learners. *International Journal of Humanities and Social Science* 7(9), 125-130.
- Bird, S.A., & Williams, J.N. (2002). The Effect of Bimodal Input on Implicit and Explicit Memory: An Investigation into the Benefits of Within-Language Subtitling. *Applied Psycholinguistics*, 23(4), 509-533. <https://doi.org/10.1017/S0142716402004022>
- Bisson, M.J., van Heuven, W.J.B., Conklin, K., & Tunney, R.J. (2015). The role of verbal and pictorial information in multimodal incidental acquisition of foreign language vocabulary. *Quarterly Journal of Experimental Psychology*, 68(7), 1306-1326. <https://doi.org/10.1080/17470218.2014.979211>
- Bright, K.S., Crane, D., & Kauffman, M. (Producers). (1995). Friends [Television series]. New York: NBC Studios.
- Caruana, S. (2021). An Overview of Audiovisual Input as a Means for Foreign Language Acquisition in Different Contexts. *Language and Speech*. <https://doi.org/10.1177/0023830920985897>
- Compleat Web VP. Retrieved March 2, 2021, from <https://www.lex tutor.ca/vp/comp/>
- Corpus of Contemporary American English. Retrieved March 2, 2021, from <https://www.english-corpora.org/>
- Danan, M. (1992). Reversed Subtitling and Dual Coding Theory: New directions for foreign language instruction. *Language Learning*, 42(4), 497- 527. <https://doi.org/10.1111/j.1467-1770.1992.tb01042.x>
- d'Ydewalle, G. (2002). The Case of Subtitling: Foreign-Language Acquisition by Watching Subtitled Television Programs. *Journal of Foreign Language Education and Research*, 12, 59-77.
- Fazilatfar, A.M., Shahin, G., & Samavarchi, L. (2011). The effect of standard and reversed subtitling versus no subtitling mode on L2 vocabulary learning. *The Journal of Teaching Language Skills*, 3(1), 43-64.
- Fievez, I., Montero Perez, 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), 21-41. <https://doi.org/10.1558/cj.39370>
- Garnier, M. (2014). Intentional vocabulary learning from watching DVDs with subtitles: A case study of an "average" learner of French. *International Journal of Research Studies in Language Learning*, 3(1), 21-32. <https://doi.org/10.5861/ijrsl.2013.521>
- Gorjian, B. (2014). The Effect of Movie Subtitling on Incidental Vocabulary Learning among EFL Learners. *International Journal of Asian Social Science*, 4(9), 1013-1026.
- Gass, S., Winke, P., Isbell, D., & Ahn, J. (2019). How captions help people learn languages: A working-memory, eye-tracking study. *Language Learning & Technology*, 23(2), 84-104.
- Hsieh, Y. (2020). Effects of Video Captioning on EFL Vocabulary Learning and Listening Comprehension. *Computer Assisted Language Learning*, 33(5-6), 567-589. <https://doi.org/10.1080/09588221.2019.1577898>
- Jaber, R.A., & Dollar, Y.K. (2019). A Meta-analysis of the Effect of Bimodal Subtitling on Vocabulary Learning among Adult EFL Learners. *Focus on ELT Journal*, 1(1), 43-52. <https://doi.org/10.14744/felt.2019.00001>
- 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>
- Koolstra, C.M., Peeters, A.L. & Spinhof, H. (2002). The Prons and Cons of Dubbing and Subtitling. *European Journal of Communication*, 17, 325-351. <https://doi.org/10.1177/0267323102017003694>
- Kuppens, A.H. (2010). Incidental Foreign Language Acquisition from Media Exposure. *Learning, Media and Technology*, 35(1), 65-85. <https://doi.org/10.1080/17439880903561876>
- Liao, S., Kruger, J., & Doherty, S. (2020). The impact of monolingual and bilingual subtitles on visual attention, cognitive load, and comprehension. *The Journal of Specialised Translation*, 33, 70-98.
- Mardani, M., & Najmabadi, A. (2016). The Effect of Movies with Different Types of Subtitles on Incidental English Vocabulary Learning of Iranian High School Students. *Iranian Journal of Research in English Language Teaching*, 3(2), 46-57.
- Markham, P.L., Peter, L.A., & McCarthy, T.J. (2001). The Effects of Native Language vs. Target Language Captions on Foreign Language Students' DVD Video Comprehension. *Foreign Language Annals*, 34(5), 439-445. <https://doi.org/10.1111/j.1944-9720.2001.tb02083.x>
- Markham, P., & Peter, L. (2003). The Influence of English Language and Spanish Language Captions on Foreign Language Listening/Reading Comprehension. *Journal of Educational Technology Systems*, 31(3), 331-341. <https://doi.org/10.2190/BHUH-420B-FE23-ALAO>

- Matielo, R., D'Ely, R.C. S.F., & Baretta, L. (2015). The Effects of Interlingual and Intralingual Subtitles on Second Language Learning/Acquisition: A State-of-the-Art Review. *Trabalhos Em Linguística Aplicada*, 54, 161-182. <https://doi.org/10.1590/0103-18134456147091>
- Matielo, R., Oliveira, R.P. de, & Baretta, L. (2018). Subtitling, Working Memory, and L2 Learning: A Correlational Study. *Revista Brasileira de Linguística Aplicada*, 18(3). <https://doi.org/10.1590/1984-6398201812773>
- Mayer, R. (2020). *Multimedia Learning* (3rd ed.). Cambridge University Press. <https://doi.org/10.1017/9781108894333.003>
- Mitterer, H., & McQueen, J.M. (2009). Foreign Subtitles Help but Native-Language Subtitles Harm Foreign Speech Perception. *PLoS ONE*, 4(11), 4-8. <https://doi.org/10.1371/journal.pone.0007785>
- Montero Perez, M., Peters, E., & Desmet, P. (2015). Enhancing Vocabulary Learning Through Captioned Video: An Eye-Tracking Study. *Modern Language Journal*, 99(2), 308-328. <https://doi.org/10.1111/modl.12215>
- Montero Perez, M., Van Den Noortgate, W., & Desmet, P. (2013). Captioned Video for L2 Listening and Vocabulary Learning: A Meta-analysis. *System*, 41(3), 720-739. <https://doi.org/10.1016/j.system.2013.07.013>
- Muñoz, C. (2017). The Role of Age and Proficiency in Subtitle Reading: an eye-tracking study. *System*, 67, 77-86. <https://doi.org/10.1016/j.system.2017.04.015>
- Nasab, M.S. B., & Motlagh, S.F. P. (2017). Vocabulary Learning Promotion through English Subtitled Cartoons. *Communication and Linguistics Studies*, 3(1-1), 1-7.
- Pattemore, A., & Muñoz, C. (2020). Learning L2 Constructions from Captioned Audio-visual Exposure: The effect of learner-related factors. *System*, 93, 1-13. <https://doi.org/10.1016/j.system.2020.102303>
- Pattemore, A., Suárez, M.M., & Muñoz, C. (2020). Exploring L2 TV mode preferences and perceptions of learning. In CALL for widening participation: short papers from EUROCALL 2020, 272-278. Research-publishing.net. <https://doi.org/10.14705/rpnet.2020.48.1200>
- Pemberton, L., Fallahkhair, S., & J. Masthoff (2005). Learner-Centred Development of a Mobile and iTV Language Learning Support System. *Educational Technology & Society*, 8(4), 52-63.
- Peters, E., Heynen, E., & Puimège, E. (2016). Learning Vocabulary through Audiovisual Input: The differential effect of L1 subtitles and captions. *System*, 63, 134-148. <https://doi.org/10.1016/j.system.2016.10.002>
- Puimège, E., & Peters, E. (2019). Learning L2 vocabulary from audiovisual input: an exploratory study into incidental learning of single words and formulaic sequences. *The Language Learning Journal*, 47(4). <https://doi.org/10.1080/09571736.2019.1638630>
- Pujadas, G., & Muñoz, C. (2019). Extensive Viewing of Captioned and Subtitled TV Series: A study of L2 vocabulary learning by adolescents. *Language Learning Journal*, 47(4), 479-496. <https://doi.org/10.1080/09571736.2019.1616806>
- Pujadas, G., & Muñoz, C. (2020). Examining Adolescent EFL Learners' TV Viewing Comprehension through Captions and Subtitles. *Studies in Second Language Acquisition*, 42(3), 551-575. <https://doi.org/10.1017/S0272263120000042>
- Rodgers, M.P. H., & Webb, S. (2017). The effects of captions on EFL learners' comprehension of English-language television programs. *CALICO Journal*, 34(1), 20-38. <https://doi.org/10.1558/cj.29522>
- Sadiku, A. (2018). The Role of Subtitled Movies on Students' Vocabulary Development. *International Journal of Sciences: Basic and Applied Research*, 42(1), 212-221.
- Sockett, G. (2014). *The Online Informal Learning of English*. Basingstoke: Palgrave Macmillan. <https://doi.org/10.1057/9781137414885>
- Suárez, M.M., & Gesa, F. (2019). Learning vocabulary with the support of sustained exposure to captioned video: do proficiency and aptitude make a difference? *The Language Learning Journal*, 47(4). <https://doi.org/10.1080/09571736.2019.1617768>
- Sydorenko, T. (2010). Modality of input and vocabulary acquisition. *Language Learning & Technology*, 14(2), 50-73.
- Teng, F. (2019). Incidental vocabulary learning for primary school students: the effects of L2 caption type and word exposure frequency. *The Australian Educational Researcher*, 46(1), 113-136. <https://doi.org/10.1007/s13384-018-0279-6>
- Vanderplank, R. (2016). 'Effects of' and 'Effects with' Captions: How exactly does watching a TV programme with same-language subtitles make a difference to language learners? *Language Teaching*, 49(2), 235-250. <https://doi.org/10.1017/S0261444813000207>
- Vanderplank, R. (2019). 'Gist Watching Can Only Take You So Far': Attitudes, strategies and changes in behaviour in watching films with captions. *Language Learning Journal*, 47(4), 407-423. <https://doi.org/10.1080/09571736.2019.1610033>
- Vulchanova, M., Aurstad, L.M. G., Kvitnes, I.E. N., & Eshuis, H. (2015). As naturalistic as it gets: subtitles in the English classroom in Norway. *Frontiers in Psychology*, 5. <https://doi.org/10.3389/fpsyg.2014.01510>

- Wang, Y. (2019). Effects of L1/L2 Captioned TV Programs on Students' Vocabulary Learning and Comprehension. *CALICO Journal*, 36(3), 204-224. <https://doi.org/10.1558/cj.36268>
- Webb, S., & Rodgers, M.P.H. (2009a). The Lexical Coverage of Movies. *Applied Linguistics*, 30(3), 407-427. <https://doi.org/10.1093/applin/amp010>
- Webb, S., & Rodgers, M.P.H. (2009b). Vocabulary Demands of Television Programs. *Language Learning*, 59(2), 335-366. <https://doi.org/10.1111/j.1467-9922.2009.00509.x>
- Wisniewska, N., & Mora, J.C. (2020). Can Captioned Video Benefit Second Language Pronunciation? *Studies in Second Language Acquisition*, 42(3). <https://doi.org/10.1017/S0272263120000029>
- Winke, P., Gass, S., & Sydorenko, T. (2010). The Effects of Captioning Videos Used for Foreign Language Listening Activities. *Language Learning & Technology*, 14(1), 65-86.
- Yang, J. (2018). Use of L1 and L2 Subtitles for Incidental Vocabulary Learning in a Second Language: Focusing on Word Frequency. *STEM Journal*, 19, 127-147. <https://doi.org/10.16875/stem.2018.19.3.127>

APPENDIX A: Target language vocabulary

Target term	Occurrences in video clips	Lexical band ¹	Lexical frequency ²
<i>Subtitled³</i>			
Steep	1	K4	190,626
Jerk	1	K4	52,576
Weird	2	K2	415,206
Prawns	1	K9	21,678
Issue	1	K1	5,208,373
<i>Captioned⁴</i>			
Tough	1	K2	863,626
Trapped	1	K2	17,897
Cabin	2	K4	326,656
Audition	2	K6	69,048
Sober	2	K5	47,189
Hardware store	2	K4	23,474
<i>Reverse subtitled⁵</i>			
Tilt	3	K4	56,203
Buzzer	1	K4	16,039
Dull	1	K4	122,138
Smitten	2	K10	15,081
Squeeze	1	K3	203,860
Dessert	1	K5	224,535

¹ Compleat Web VP, 2021.

² Corpus of Contemporary American English, 2021.

³ Brown, C. (Writer) & Gittelsohn, E. (Director). (19 October 1995). The One with Five Steaks and an Eggplant. Episode 5 In Bright/Kauffman/Crane Productions & Warner Bros. Television, Friends. New York: NBC Studios.

⁴ Ungerleider, I. (Writer) & Schlamme, T. (Director). (4 January 1996). The One with Russ. Episode 10 In Bright/Kauffman/Crane Productions & Warner Bros. Television, Friends. New York: NBC Studios.

⁵ Curtis, M., Malins, G.S. (Writers), & Lembeck, M. (Director). (8 February 1996). The One Where Ross and Rachel... You Know. Episode 15 In Bright/Kauffman/Crane Productions & Warner Bros. Television, Friends. New York: NBC Studios.

