ELIA
Estudios de Lingüística Inglesa Aplicada

THE EFFECTS OF REGULAR AND ENHANCED CAPTIONS ON INCIDENTAL VOCABULARY ACQUISITION

EFECTOS DE LA SUBTITULACIÓN EN L2 Y DEL REALCE TEXTUAL EN LA ADQUISICIÓN INCIDENTAL DE VOCABULARIO

Rebeca Finger-Bou
Universitat de Barcelona
rfingerb@ub.edu

Carmen Muñoz
Universitat de Barcelona
munoz@ub.edu

Abstract

Research has demonstrated that watching audiovisual materials in the target language (L2) through using captions can foster vocabulary learning. Some studies have redirected learners' attention by enhancing specific parts of those captions, thus increasing their saliency. This study explores the effects of regular and enhanced captions on incidental vocabulary acquisition by L1-Spanish/Catalan learners of English through short exposure to a documentary. It also analyses how vocabulary learning might be affected by previous vocabulary knowledge and language learning aptitude. Two randomly distributed groups were formed. One was provided with

regular captions, and the other with enhanced captions (target words in yellow and bold). Vocabulary gains were assessed through pre- and post-tests that tapped into meaning recall, meaning recognition and form recognition knowledge. The results showed that the difference between the pre-test and the post-test was greater in the students with enhanced subtitles, but the difference was not significant between the two groups in the post-test. Vocabulary size emerged as the most significant predictor, but not aptitude. Retrospective questionnaires on participants' focus of attention reported an emphasis on captions and comprehension. Analyses indicate that paying attention to the enhanced items might have positively affected acquisition and retention. This study has provided new evidence that shows the potential advantage of multimodal input as an accessible pedagogical tool for acquiring languages.

Keywords: multimodal input, incidental vocabulary acquisition, enhanced captions, regular captions, individual differences, focus of attention.

Resumen

Un buen número de estudios ha demostrado que el visionado de material audiovisual en la lengua meta (L2) con la ayuda de subtítulos en la L2 facilita el aprendizaje de vocabulario. Algunos de ellos han dirigido la atención de los participantes realzando partes específicas de esos subtítulos, aumentando así su prominencia. Este estudio explora los efectos de los subtítulos normales y realzados en la adquisición incidental de vocabulario por parte de estudiantes de inglés cuyas lenguas maternas son el español y el catalán, a través de una breve exposición a un documental. También analiza cómo el aprendizaje de vocabulario puede verse afectado por el conocimiento previo de vocabulario y la aptitud para aprender idiomas. Se formaron dos grupos aleatoriamente. Uno visionó el documental con subtítulos normales en inglés y, el otro, con subtítulos realzados (palabras clave en amarillo y negrita). El aprendizaje de vocabulario

se evaluó mediante pruebas previas y posteriores de recordar el significado y la forma, y de reconocer el significado. Los resultados mostraron que la diferencia entre las pruebas era mayor en los estudiantes con subtítulos realzados, pero la diferencia no era significativa entre los dos grupos en la prueba posterior. El tamaño de vocabulario previo resultó ser la variable explicativa más significativa, pero no así la aptitud. Los cuestionarios retrospectivos revelaron atención en los subtítulos y la comprensión. Los análisis indican que prestar atención a los elementos realzados podría haber afectado positivamente a la adquisición y posterior retención de vocabulario. Este estudio proporciona nueva evidencia sobre el beneficio potencial del input multimodal como una herramienta pedagógica accesible para la adquisición de idiomas.

Palabras clave: input multimodal, adquisición incidental de vocabulario, subtítulos, subtítulos realzados, diferencias individuales, foco de atención.

1. Literature Review

The massive access to multimodal second language (L2) input in modern times is one of the major reasons that may explain the growing importance of multimodal input in the area of Second Language Acquisition (SLA) (Montero Perez, 2022). Several studies have revealed that language learners are indeed motivated to watch television in an L2 (Peters & Muñoz, 2020). Mayer's (2014) cognitive theory of multimedia learning ascertains that language learning is greater when information is not only processed in spoken mode but also in written mode, for learners produce mental connections between the aural and the visual information, providing that there is a temporal proximity. In that sense, television programs also supply L2 learners with repeated encounters with both high-frequency and low-frequency words (Rodgers & Webb, 2011), which could potentially fuel L2 vocabulary growth with regular viewing (Feng & Webb, 2019).

Considering multimodal input as a combination of pictorial information, written verbal information –in the form of captions or subtitles–, and acoustic verbal input (Peters & Muñoz, 2020), multimodal input can enhance language learning whenever all channels, that is, visual and verbal information, are activated simultaneously (Montero Perez, 2022).

1.1. Vocabulary Acquisition through Multimodal Input

Based on Paivio's (1986) dual coding theory, Mayer's (2014) model of multimedia learning proposes that learning is more effective with both words and pictures compared to when words or pictures alone are present. In that sense, research has sought to throw light upon the effects of captions on language acquisition in general and has succeeded in doing so by demonstrating the statistically significant advantage of participants who watch multimodal materials with captions (subtitles in the L2 or original language) or subtitles (in the L1). The use of captions has hence been corroborated to have a positive impact on L2 comprehension (Montero Perez et al., 2014), vocabulary (Pujadas & Muñoz, 2019; Suárez & Gesa, 2019) and grammar (Pattemore & Muñoz, 2020) learning, as word recognition is assisted by the breaking down of speech into separated items.

1.2. Enhanced Captioning

Ample evidence exists confirming vocabulary acquisition through multimodal input and the use of captions (Montero Perez, 2022; Muñoz, 2022). Furthermore, recent studies have intended to redirect and refocus learners' attention by typographically enhancing specific parts of those captions (Majuddin et al., 2021; Montero Perez et al., 2014; Lee & Révészc, 2020), as noticing has been widely recognised as

a relevant and essential part of language learning (Schmidt, 1994). When the material salience of single-words is typographically enhanced in captions, learners of an L2 will expectedly pay more attention and learn new L2 vocabulary items (Montero Perez et al., 2015; Puimège et al., 2022). However, research has not revealed a clear advantage of textually enhanced captions over regular or unenhanced captions, and the different techniques that have been used (e.g., highlighting, bolding or using keywords) have shown mixed results (Montero Perez, 2022).

1.3. Individual Differences

In this line of research, it has been found that learner-related factors such as proficiency level (Montero Perez et al., 2013; Suárez & Gesa, 2019), previous vocabulary knowledge (Majuddin, 2020; Montero Perez et al., 2014; Peters & Webb, 2018; Rodgers & Webb, 2019), or working memory (Pattemore & Muñoz, 2020) may impact vocabulary gains and the processing of multimodal input. Particularly, prior vocabulary knowledge is one of the most important factors affecting incidental vocabulary acquisition (Peters & Webb, 2018).

In contrast, only a few studies have analysed the association of language learning aptitude with vocabulary learning through multimodal input. In Suárez and Gesa's (2019) study, for example, aptitude was found to be statistically significant only in the learning of target word meanings, not forms, after exposure to captioned videos. Moreover, the authors also found a main effect for proficiency on the learning scores for both target word forms and meanings. Contrary to the mentioned study, however, Pattemore and Muñoz (2020) did not find any significant effect of the LLAMA tests on grammar construction learning from captioned audio-visual exposure. The authors propose that learners might cease to rely on language learning aptitude when surpassing a certain proficiency threshold, as suggested by Winke (2013). However, previous studies

have not used enhanced captions and it is still unknown if vocabulary size and aptitude play a similar role as when captions are unenhanced.

1.4. Learners' Focus of Attention

Learners' focus of attention while viewing captioned material needs to be investigated. This may be done through retrospective questionnaires that resemble think-aloud verbal protocols (Winke, 2013), which allows researchers to extract subjective and self-reflective information on the conducted experiments. With these conditions in mind, the present study will aim at analysing the effects of caption enhancement on incidental vocabulary acquisition in L1-Spanish/Catalan students of English as a Foreign Language. More specifically, this study pursues to answer the following research questions:

1.5. Research Questions

1. Is there evidence of incidental vocabulary acquisition after viewing a captioned documentary? If so, is the potential learning retained after two weeks?
2. Does the enhancement of captions have an effect on incidental vocabulary acquisition in comparison to regular captions in L1-Spanish/Catalan EFL learners? If so, is the potential learning retained after two weeks?
3. To what extent do previous vocabulary knowledge and language learner's aptitude, as measured by LLAMA B and D, play a role in potential vocabulary gains through viewing a captioned documentary?

4. How do enhanced and regular captions affect L1-Spanish/Catalan EFL learners' self-reported focus of attention when viewing a captioned documentary?

2. Methodology

2.1. Participants

The participants of this study consisted of 31 L1-Spanish/Catalan learners of English, who were enrolled in different EFL levels at a language school in a small city in the province of Tarragona, Spain. The participants included 18 adolescents and 13 adults, and ages varied from 14 to 64 years old (M = 22.46, SD = 11.11). All participants had the same teacher, who collaborated with the researchers.

A background information questionnaire was handed out prior to the experiment so that personal information such as age, sex and previous education could be collected, as well as information on external sources of input, that is, out-of-school exposure to L2 media. Parental consent forms were distributed to all underaged students, whereas adult learners signed to accept their own participation. Two randomly distributed groups were formed. Group 1 was provided with regular captions (RC), and group 2 with enhanced captions (EC).

Table 1: *Descriptive information of participants*

	Age				Level				Sex
	Mean	SD	Min	Max	B1	B1+	B2	C1	
Regular (n = 15)	19.65	6.03	14.00	37.60	2	5	5	3	6 female, 9 male
Enhanced (n = 16)	25.09	14.06	15.00	64.00	4	2	6	4	10 female, 6 male
All participants (n = 31)	22.46	11.11	14.00	64.00	6	7	11	7	16 female, 15 male

2.2. Target Constructions

A total of 21 target words (TWs) were chosen from the script of *Viral: The 5G Conspiracy Theory* (Livingston, 2020), a 25-minute documentary from the BBC that was released in 2020 in which the conspiracy theories that erupted ever since the beginning of the global pandemic are critically reviewed. Words in the documentary were assessed through *LexTutor* to extract 21 TWs (from the 1k, 2k, Academic Word List, and OFF types) that appeared at least twice in the audio-visual material. The TWs included in the final analysis were *spread, lockdown, harmful, jaw, mad, murderer, opposed, threat, approach, linked, network, remove, appealing, arson, carer, cell, clap, lockdown, mast, ripper,* and *illness*. An enhanced version of the regular captions was created with the application *SubtitleEdit* (v3.5.18) and embedded on the video with *HandBrake* (v1.3.0-v1.3.3), where TWs were presented in yellow and bold. Moreover, a virtually equivalent number of words that belonged to the same frequency lists and did not appear in the documentary were selected to function as distractors. TWs and distractors were revised and approved by the participants' teacher.

2.3. Instruments

Vocabulary gains were assessed through pre-, immediate post- and delayed post-tests that tapped into meaning knowledge at the level of recall and recognition, to gather information at the two different sensitivities based on Nation's (2001) nine components of word knowledge. Additionally, immediate post- and delayed post-tests on form recognition were included to assess whether learners remembered seeing TWs on the documentary, as noticing a new word is the first step towards acquisition and it has been suggested that captions generally help learners with both written and aural form recognition and with developing form-meaning connections (Pujadas

& Muñoz, 2020). When taking the tests, TWs and distractors were provided through an audio file recorded with the teacher's voice that repeated each word twice, whilst the written forms could be read in the paper where participants were to answer, which guaranteed them encountering the same modalities in the tests as those in the multimodal input, and therefore all channels of input were re-activated simultaneously. All tests were piloted by five L1-Spanish/Catalan learners of English whose ages (M = 36.8, SD = 14.9) ranged very similarly to the study's participants (M = 22.46, SD = 11.11). Pre-test scores in the two languages were similar (SPA (n = 24) = 72.4% vs CAT (n = 7) = 70.7%)[1].

Participants' previous vocabulary knowledge was measured by means of Meara and Miralpeix's (2015) *V_YesNo* (v1.01) test. Language learning aptitudes for vocabulary learning and listening for new words, the most relevant to this study, were measured through two of the subtests of Meara and Rogers's (2019) LLAMA suite of tests: *LLAMA B* (v3.00) and *D* (v3.00). LLAMA B consists of a vocabulary learning task in which participants must remember large amounts of words. This subtest measures the users' ability to attach unfamiliar names to unfamiliar objects LLAMA D, on the other hand, is a phonetic memory subtest, where users must recognise spoken language that they were exposed to a short while earlier (Rogers et al., 2017).

An additional test with three comprehension questions that had no relation to the TWs was utilised in the immediate post-test. Furthermore, Likert-scale questionnaires resembling think-aloud verbal protocols (Winke, 2013) were also distributed so as to collect retrospective information on learners' self-reported focus of attention.

[1] A series of chi-squared tests revealed that only for one item (*ripper*) the difference between difficulty indexes in the two languages was statistically significant (p = .011). Perceptions reflected in the retrospective questionnaire did not account for any extra difficulty when dealing with different items in the meaning recognition tests.

2.4. Procedure

The five experimental sessions were organised during regular class time across three consecutive months between the second and third trimesters of the academic year (Table 2). The nature of the experiment was unknown to all participants and the teacher did not provide any extra practice on vocabulary.

Table 2: Experimental procedure

Session 1	Session 2	Session 3	Session 4	Session 5
Background information questionnaire	Pre-test	Individual differences	Documentary viewing + Immediate post-test	Delayed post-test
Out-of-school exposure to L2 media	Meaning recall	V_YesNo	Comprehension test (T/F)	Form recognition + meaning recall
Consent form	Meaning recognition	LLAMA B	Form recognition + meaning recall	Meaning recognition
		LLAMA D	Meaning recognition	Retrospective questionnaire

During the first two weeks, participants completed the background questionnaire, the vocabulary size test, the language learning aptitude tests, and the pre-test. Six weeks later, all subjects watched the documentary with either regular or enhanced captions, and then immediately answered three true or false comprehension questions that had no relation to the TWs, as well as post-tests on form recognition, meaning recall and meaning recognition. That is, students were asked whether they had seen a particular item in the documentary, whether they could provide a translation for that item, and whether they could identify the correct translation of the item out of four options. Two weeks later, a delayed post-test was carried out to compare the effectiveness of these treatments in the short- and the long-term. Finally, participants completed the retrospective questionnaire on learners' self-reported focus of attention, to identify

their reactions and emphasis when conducting the study according to their own perceptions.

Due to the pandemic, all classes over six students had to be conducted online until mid-May 2021. For that reason, sessions 1 to 4 were performed online for the eight students from 2nd of *Batxillerat* (B2) and the seven adults attending the Advanced class (C1). All materials and procedures were transposed to an online environment (*Google Forms* for the tests and *Edpuzzle* for the viewing of the documentary) to imitate as accurately as possible the in-person format of the experiment. In that regard, tests and viewing sessions were undertaken during class time but in an online environment. The rest of participants (16 in total) were able to complete all tasks face-to-face from beginning to end.

2.5. Scoring Data and Analysis

One point was assigned for a right answer per item. A mean for all answers was estimated, for a total of 1 point per test, which was then multiplied by 100 in the reports, to aid visualise and understand group differences.

The normal distribution of all groups' scores was assessed and confirmed through the software *IBM SPPS Statistics 25 version*. Several independent samples *t*-tests were conducted to assess the comparability between experimental groups. Individual differences such as vocabulary size scores ($p = .338$), LLAMA B scores ($p = .349$) and LLAMA D scores ($p = .384$) between the two groups were normally distributed and non-significantly different[2]. Next, a series of Generalized Linear Mixed Models (GLMMs) were used to answer the

[2] Pre-test analyses showed that independent variables were not highly correlated ($r < .7$), and that independent and dependent variables were significantly related ($p < .05$). Only correlations between LLAMA B and form recognition scores were found non-significant ($p = .151$).

first, second and third research question. Furthermore, quantitative and qualitative analyses of the retrospective questionnaires were undertaken to understand learners' self-reported focus of attention when watching the documentary with regular or enhanced captions.

3. Results

The descriptive statistics of the variables of the two groups, as well as those for all participants, appear in Tables 3 and 4.

Table 3: Descriptive statistics: Individual differences

	Vocabulary size (max: 10000)				LLAMA B (max: 20)				LLAMA D (max: 20)			
	Mean	SD	Min	Max	Mean	SD	Min	Max	Mean	SD	Min	Max
Regular (n = 15)	4917.80	1053.16	3096	6537	11.13	4.14	5	20	9.33	4.12	2	15
Enhanced (n = 16)	5343.38	1347.43	3310	7704	9.63	4.65	3	20	8.19	3.06	2	13
All participants (n = 31)	5137.45	1213.31	3096	7704	10.35	4.40	3	20	8.74	3.60	2	15

A series of independent *t*-tests showed that there were no significant differences between their pre-test scores at meaning recall ($p = .248$) or meaning recognition ($p = .870$), even though the mean score of the EC group was always slightly higher than the RC group. Moreover, an additional analysis of the comprehension task revealed that students responded correctly more than 90% of the time, and there was no significant difference in comprehension between the two experimental groups ($p = .654$).

Analyses were conducted separately for meaning recall, meaning recognition and form recognition using GLMMs. Neither *LLAMA B* nor *LLAMA D* scores had significant main effects, and thus they were eliminated from the final models. Similarly, another

Table 4: Descriptive statistics: Tests

	Meaning recall											
	Pre-test score (max: 21)				Post-test score (max: 21)				Delayed post-test score (max: 21)			
	Mean	SD	Min	Max	Mean	SD	Min	Max	Mean	SD	Min	Max
Regular (n = 15)	10.80	5.65	1.00	18.00	12.07	5.52	1.00	19.00	12.73	5.41	2.00	21.00
Enhanced (n = 16)	11.75	4.63	3.00	19.00	14.00	5.24	3.00	20.00	13.37	4.79	4.00	21.00
All participants (n = 31)	11.29	5.08	1.00	19.00	13.06	5.38	1.00	20.00	13.06	5.02	2.00	21.00
	Meaning recognition											
	Pre-test score (max: 21)				Post-test score (max: 21)				Delayed post-test score (max: 21)			
	Mean	SD	Min	Max	Mean	SD	Min	Max	Mean	SD	Min	Max
Regular (n = 15)	15.07	3.53	10.00	20.00	16.20	4.36	8.00	21.00	15.73	4.62	8.00	21.00
Enhanced (n = 16)	15.19	3.73	8.00	20.00	17.56	3.86	9.00	21.00	17.06	3.45	9.00	21.00
All participants (n = 31)	15.13	3.58	8.00	20.00	16.90	4.10	8.00	21.00	16.42	4.05	8.00	21.00
	Form recognition											
	Post-test score (max: 21)				Delayed post-test score (max: 21)				Comprehension test (max: 3)			
	Mean	SD	Min	Max	Mean	SD	Min	Max	Mean	SD	Min	Max
Regular (n = 15)	14.86	1.72	11.93	18.61	12.98	2.00	10.02	16.23	2.80	0.56	1	3
Enhanced (n = 16)	15.72	2.60	10.02	19.09	13.33	2.72	7.16	18.61	2.88	0.34	2	3
All participants (n = 31)	15.30	2.22	10.02	19.09	13.16	2.36	7.16	18.61	2.84	0.45	1	3

explored fixed factor was *Level*, that is, the level in which participants were enrolled in at the school (B1, B1+, B2 and C1). As only in one of the following analyses *Level* was found significant, it was eliminated from every other model. However, vocabulary size scores did have a significant main effect in all tests. For that reason, the common fixed factors in all the remaining models were *Vocabulary Size* alongside with *Time* (pre-test, immediate post-test, and delayed post-test), whereas *Subject* (participants) and *Item* (TWs) were included as random intercepts.

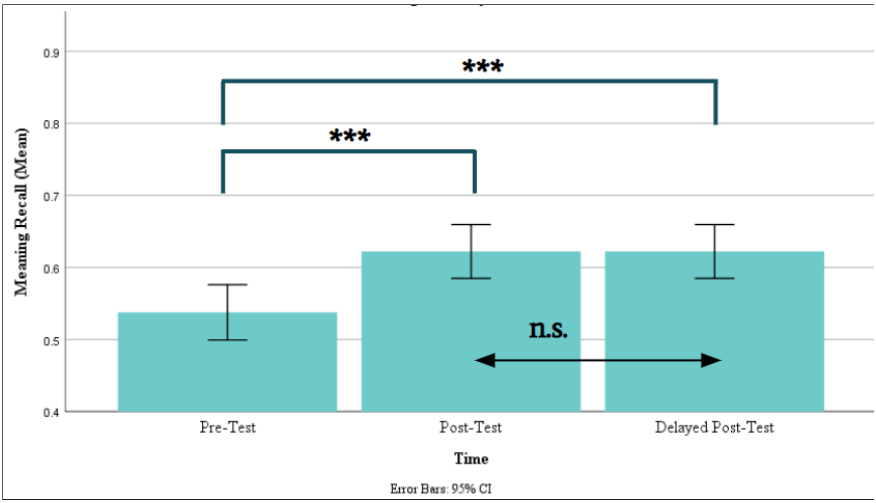
3.1. Vocabulary Acquisition

Meaning Recall

In relation to the first research question, both groups showed improvement from pre-test to both post-tests in meaning recall (Figure 1). A significant interaction between the level in which participants were enrolled and time at testing was found ($p = .034$), and thus *Level* was included as another fixed effect in the GLMM only for this variable. Pairwise comparisons of scores at pre-test, immediate post-test and delayed post-test showed that differences between pre-test and immediate post-test and pre-test and delayed post-test were significant ($p < .001$ in both cases), whereas differences between immediate post-test and delayed post-test were not ($p = .425$) (see Figure 1).

Significant main effects of *Vocabulary Size* (F (1,1940) = 17.117, p <.001), *Time* (F (2, 1940) = 13.584, $p < .001$) and a significant interaction between *Level* and *Time* (F (6, 1940) = 2.277, $p = .034$) were found in the analysis, as well as a non-significant main effect of Level (F (3, 1940) = 1.082, $p = .356$). The Bonferroni adjusted results revealed that all groups improved significantly from pre-test to immediate post-test ($p = .002$) (Table 5).

Figure 1: Meaning recall by time



n.s. → $p > .05$
** → $p < .05$
*** → $p < .001$

Table 5: Results from GLMM: fixed coefficients for meaning recall regardless of condition

	Coefficient	SE	t	Sig.	95% Confidence Interval		Exp (Coefficient)	95% CI for Exp (Coefficient)	
					Lower	Upper		Lower	Upper
Intercept	-3.404	1.520	-2.240	.025	-6.385	-.423	.033	.002	.655
V_Size	.001	< .001	4.137	**< .001**	< .001	.001	1.001	1.000	1.001
Time=1	-1.083	.345	-3.140	**.002**	-1.760	-.407	.338	.172	.666
Time=2	-.067	.367	-.184	.854	-.788	.653	.935	.455	1.921
Level=1	-.206	.778	-.265	.791	-1.733	1.320	.814	.177	3.745
Level=2	-.896	.801	-1.118	.264	-2.467	.675	.408	.085	1.964
Level=3	-.428	.714	-.599	.549	-1.828	.973	.652	.161	2.645
[Level=1] * [Time=1]	-.113	.492	-.230	.818	-1.077	.851	.893	.341	2.342

[Level=2] * [Time=1]	.300	.482	.623	.533	-.645	1.245	1.350	.525	3.472
[Level=3] * [Time=1]	1.029	.416	2.474	**.013**	.213	1.845	2.799	1.238	6.327
[Level=1] * [Time=2]	-.535	.506	-1.057	.290	-1.529	.458	.585	.217	1.581
[Level=2] * [Time=2]	-.262	.495	-.529	.597	-1.233	.709	.769	.291	2.031
[Level=3] * [Time=2]	.544	.438	1.244	.214	-.314	1.402	1.723	.731	4.065

More specifically, pairwise comparisons between *Level* and *Time* show significant differences, on the one hand, between pre-test and delayed post-test only for B1 (p = .003) and C1 (p = .018), with B1+ nearly significant (p = .052) and B2 differences non-significant at all (p = .816). On the other hand, only the C1 level managed to show significant differences between pre- and immediate post-tests as well (p = .018).

Meaning Recognition

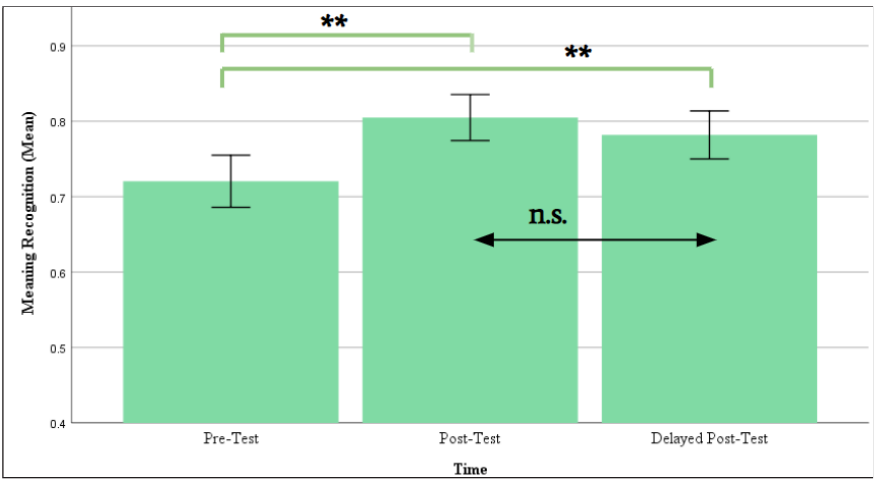
At the meaning recognition level, the GLMM showed a significant main effect of *Vocabulary Size* (F (1,1947) = 32.152, p < .001), alongside a non-significant main effect of *Time* (F (2, 1947) = 1.303, p = .272) and a significant interaction of *Vocabulary Size* and *Time* (F (2,1947) = 3.472, p = .031) (see Table 6).

Pairwise comparisons showed significant differences between pre- and immediate post-test as well as pre- and delayed post-test (p = .014 and p = .021, respectively). A non-significant difference between immediate post- and delayed post-test (p = .226) further suggests that word knowledge was not significantly lost (see Figure 3).

Table 6: Results from the GLMM: fixed coefficients for meaning recognition regardless of experimental group

	Coefficient	SE	t	Sig.	95% Confidence Interval		Exp (Coefficient)	95% CI for Exp (Coefficient)	
					Lower	Upper		Lower	Upper
V_Size	.001	.000	5.286	**< .001**	.001	.001	1.001	1.001	1.001
Time=1	1.036	.773	1.339	.181	-.481	2.552	2.817	.618	12.833
Time=2	-0.076	.838	-.091	.928	-1.719	1.567	0.927	.179	4.790
V_Size * [Time=1]	< .001	< .001	-2.048	**.041**	-.001	< .001	1.000	.999	1.000
V_Size * [Time=2]	< .001	< .001	.366	.715	< .001	< .001	1.000	1.000	1.000

Figure 3: Meaning recognition by time



n.s. → $p > .05$
** → $p < .05$
*** → $p < .001$

Form Recognition

As summarized in Table 7, the GLMM revealed a significant main effect of *Vocabulary Size* (F (1, 1298) = 15.439, p < .001), a non-significant main effect of *Time* (F (1, 1298) = 3.208, p = .074) and a non-significant interaction between these two fixed factors (p = .516) for form recognition results.

Table 7: Results from the GLMM: fixed coefficients for form recognition regardless of experimental group

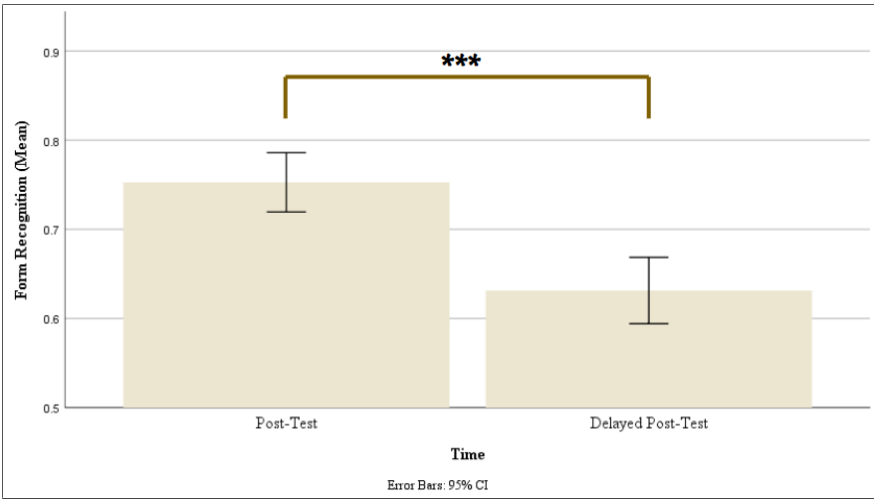
	Coefficient	SE	t	Sig.	95% Confidence Interval		Exp (Coefficient)	95% CI for Exp (Coefficient)	
					Lower	Upper		Lower	Upper
V_Size	< .001	< .001	3.881	**< .001**	< .001	.001	1.000	1.000	1.001
Time=2	1.078	.602	1.791	.074	-.103	2.258	2.938	.902	9.568
V_Size* [Time=2]	< .001	< .001	-.650	.516	< .001	< .001	1.000	1.000	1.000

Form recognition differences between immediate post-and delayed post-tests showed a significant reduction of accuracy (p < .001) from one time to the other (Figure 4).

3.2. The Effects of Enhanced Captions

The second research question focused on the effects of enhanced captions on L2 vocabulary acquisition in comparison to regular captions. As has been seen in Table 4, all participants gained knowledge no matter their proficiency level, as differences between pre-, immediate post- and delayed post-tests showed a general and significant increase in scores. GLMMs were used to estimate differences between experimental groups. After analysing which independent variables had significant effects, only *Vocabulary Size* and *Time* were maintained as fixed factors and *Subject* and *Item* as

Figure 4: (Target) Form recognition by time



n.s. → $p > .05$
** → $p < .05$
*** → $p < .001$

random intercepts, with *Caption* (regular, enhanced) as the new included fixed factor.

Meaning Recall

The GLMM for meaning recall showed significant main effects of *Vocabulary Size* (F (1, 1946) = 27.692, $p < .001$) and *Time* (F (2, 1946) = 11.067, $p < .001$), with non-significant effects of *Caption* and the interaction between *Caption* and *Time* (p =.759 and p = .289, respectively) (see Table 8).

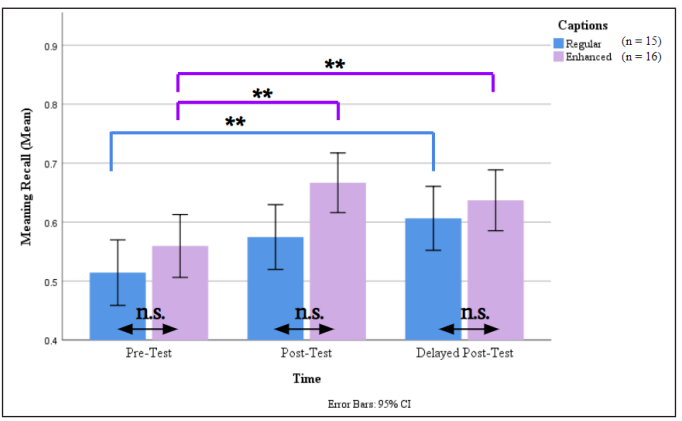
Even though comparisons at specific testing times were non-significant between the two groups (p = .898, p = .394 and p = .890 for pre-, post- and delayed post-test), the pairwise contrasts showed significant differences between both pre- and immediate post-test (p =

Table 8: Results from the GLMM: fixed coefficients of meaning recall with caption distinction

	Coefficient	SE	t	Sig.	95% Confidence Interval		Exp (Coefficient)	95% CI for Exp (Coefficient)	
					Lower	Upper		Lower	Upper
V_Size	.001	< .001	5.262	**< .001**	.001	.001	1.001	1.001	1.001
Caption=1	.064	.465	.138	.890	-.847	.975	1.066	.429	2.651
Time=1	-.556	.208	-2.675	**.008**	-.964	-.148	.573	.381	.862
Time=2	.224	.212	1.058	.290	-.191	.640	1.251	.826	1.896
[Caption=1] * [Time=1]	-.123	.301	-.411	.681	-.713	.466	.884	.490	1.594
[Caption=1] * Time=2]	-.463	.305	-1.520	.129	-1.061	.135	.629	.346	1.144

.002) and pre- and delayed post-test (p = .020) in the enhanced captions group, whereas the regular captions group only showed significant differences between pre- and delayed post-tests (p = .009) (Figure 5).

Figure 5: Meaning recall by time by captions



n.s. → $p > .05$
** → $p < .05$
*** → $p < .001$

Meaning Recognition

For meaning recognition, the GLMM showed a significant main effect of *Vocabulary Size* (F (1, 1941) = 30.718, p < .001), as well as a marginally significant interaction between *Vocabulary Size* and *Time* (F (2, 1941) = 2.916, p = .054) (see Table 9).

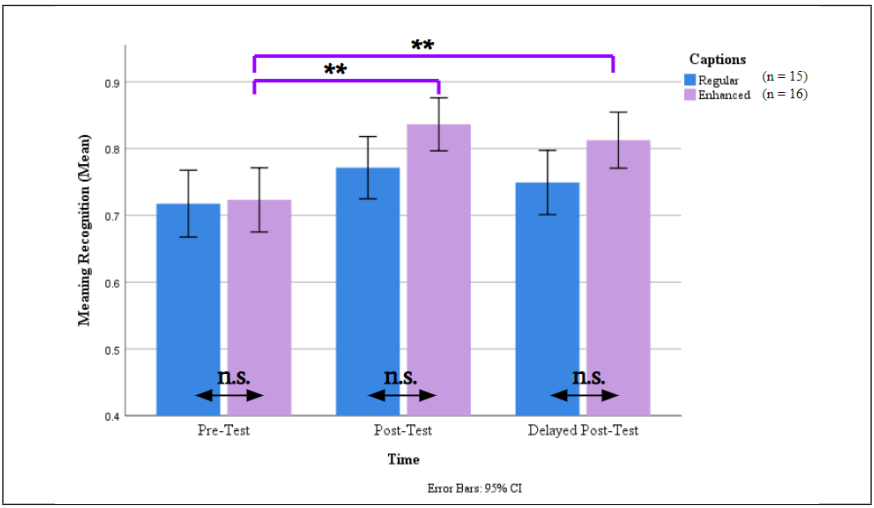
Table 9: Results of the GLMM: fixed coefficients for meaning recognition with caption distinction

	Coefficient	SE	t	Sig.	95% Confidence Interval		Exp (Coefficient)	95% CI for Exp (Coefficient)	
					Lower	Upper		Lower	Upper
V_Size	.001	< .001	3.496	**< .001**	< .001	.001	1.001	1.000	1.001
Caption=1	-1.359	2.063	-0.659	.510	-5.405	2.687	.257	.004	14.690
Time=1	.937	1.110	.845	.398	-1.239	3.114	2.553	.290	22.500
Time=2	-.415	1.258	-.330	.741	-2.882	2.052	.660	.056	7.782
V_Size* [Time=1]	< .001	< .001	-1.572	.116	-.001	< .001	1.000	.999	1.000
V_Size* [Time=2]	< .001	< .001	.542	.588	< .001	.001	1.000	1.000	1.001
[Caption=1] * [Time=1]	-.065	1.562	-.042	.967	-3.129	2.999	0.937	.044	20.061
[Caption=1] * [Time=2]	.630	1.701	.370	.711	-2.705	3.965	1.877	.067	52.725
V_Size * [Caption=1] *[Time=1]	< .001	< .001	.851	.395	< .001	.001	1.000	1.000	1.001
V_Size * [Caption=1] *[Time=2]	< .001	< .001	.162	.871	-.001	.001	1.000	.999	1.001
V_Size * [Caption=1] * [Time=3]	< .001	< .001	.525	.599	-.001	.001	1.000	.999	1.001

Pairwise comparisons between experimental groups displayed significant differences between testing times exclusively for the EC

group (p = .030 and p = .033 for pre- vs immediate post-test and pre- vs delayed post-test, respectively) (Figure 6). That is to say, participants who watched the documentary with enhanced captions significantly increased their score in both the immediate post-test and the delayed post-test and gained a significant amount of knowledge at the meaning recognition level, whereas participants in the regular captions group did not. RC's scores did not differ significantly between any of the three time points, even if the scores did tend to increase.

Figure 6: Meaning recognition by time by captions



n.s. → $p > .05$
** → $p < .05$
*** → $p < .001$

Form Recognition

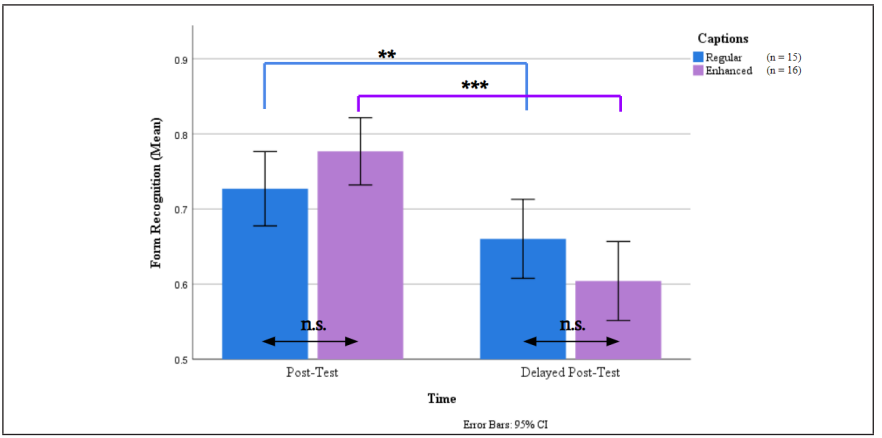
As for the form recognition test, the GLMM showed that both *Vocabulary Size* (F (1,1294) = 13.075, p < .001) and the interaction between *Caption* and *Time* (F (1,1294) = 4.707, p = .030) were statistically significant (Table 10).

Table 10: Results from the GLMM: fixed coefficients for form recognition with caption distinction

	Coefficient	SE	t	Sig.	95% Confidence Interval		Exp (Coefficient)	95% CI for Exp (Coefficient)	
					Lower	Upper		Lower	Upper
V_Size	.001	< .001	3.883	**< .001**	< .001	.001	1.001	1.000	1.001
Caption=1	2.265	1.394	1.624	.105	-.470	4.999	9.626	.625	148.300
Time=2	2.401	.828	2.900	**.004**	.777	4.025	11.033	2.174	55.981
V_Size * [Time=2]	< .001	< .001	-1.723	.085	-.001	< .001	1.000	.999	1.000
[Caption=1] * [Time=2]	-2.660	1.226	-2.170	**.030**	-5.066	-.255	.070	.006	.775
V_Size * [Caption=1] * [Time=2]	< .001	< .001	.137	.891	-.001	.001	1.000	.999	1.001
V_Size * [Caption=1] * [Time=3]	< .001	< .001	-1.366	.172	-.001	.000	1.000	.999	1.000

No significant differences were found at specific testing times (p = .537 for immediate post-test and p = .230 for delayed post-test). Both experimental groups had significant differences between immediate post- and delayed post-tests (p = .041 for the regular group and p < .001 for the enhanced group), with the enhanced group scoring higher at the immediate post-test (EC 81.12% vs RC 77.91%) but lower than the regular group at the delayed post-test (EC 61.61% vs RC 69.94%) (Figure 7). In that sense, participants from the RC were able to retain more information than those from the EC.

Figure 7: Form recognition by time by captions



n.s. → $p > .05$
** → $p < .05$
*** → $p < .001$

3.3 Learners' Awareness and Self-reported Focus of Attention

The fourth and final research question focused on EFL learners' self-reported focus of attention when viewing a captioned documentary. Through a series of retrospective questions, information on different levels of attention was gathered. First of all, as can be seen in Table 11, participants' self-reported focus of attention was very similar from one experimental group to the other. In fact, a series of independent *t*-tests revealed that none of the differences between percentages were statistically significant ($p > .05$).

Table 11: Participants' self-reported focus of attention (out of a total of 100%)

	Captions (%)				Audio (%)				Image (%)			
	Mean	SD	Min	Max	Mean	SD	Min	Max	Mean	SD	Min	Max
Regular (n = 15)	36.89	18.99	5	80	35.55	15.81	10	70	27.55	15.75	10	70
Enhanced (n = 16)	40.31	18.48	10	70	34.38	14.36	15	65	25.31	13.84	10	50
All participants (n = 31)	38.65	18.50	5	80	34.95	14.84	10	70	26.40	14.59	10	70

Standard deviations of the distribution are high, which indicates that the data is more spread out, or, in other words, that the mean is not that reliable. In general, participants reported to focus more on captions, followed by the audio and the image. Even though the tendency of the EC group is to focus more on captions, probably due to the enhancement of TWs, the difference with the RC group is non-significant. Results showed that the tendency of the RC group, but not of the EC group, was to focus more on the image and the audio.

Secondly, as for participants' self-reported linguistic focus of attention displayed in Table 12, differences among experimental groups were, once again, found non-significant ($p > .05$), so participants' distribution of percentages were statistically similar. The common ranking for all linguistic features in both experimental groups was: general comprehension, new vocabulary, pronunciation, expressions and intonation, in that order.

Table 12: Participants' self-reported linguistic focus of attention (out of a total of 100%)

	General Comprehension (%)				New vocabulary (%)				Pronunciation (%)			
	Mean	SD	Min	Max	Mean	SD	Min	Max	Mean	SD	Min	Max
Regular (n = 15)	38.87	18.78	15	80	18.87	10.52	0	40	17.60	8.53	4	35
Enhanced (n = 16)	41.88	17.88	15	80	17.50	7.75	5	30	16.09	10.12	3	40
All participants (n = 31)	40.42	18.08	15	80	18.16	9.06	0	40	16.82	9.26	3	40

	Expressions (%)				Intonation (%)			
	Mean	SD	Min	Max	Mean	SD	Min	Max
Regular (n = 15)	13.07	6.79	0	25	14.60	7.04	4	25
Enhanced (n = 16)	14.06	6.64	5	25	10.47	7.20	0	25
All participants (n = 31)	13.58	6.62	0	25	12.47	7.31	0	25

Regarding participants' self-reported amount of learning (Table 13), both experimental groups reported having learned similar amounts of knowledge (again, non-significantly different). The total mean, as well as the individual means per group, is between 2 and 3, which suggests that most of participants' self-perceptions range from *a little bit* to *quite something*, in line with the acquisition quantitatively registered.

Table 13: Participants' self-reported amount of learning

	Learning Perception			
	Mean	SD	Min	Max
Regular (n = 15)	2.40	0.63	1	3
Enhanced (n = 16)	2.38	0.50	2	3
All participants (n = 31)	2.39	0.56	1	3

Note. Likert-scale from 1–nothing (*nada*), 2–a little bit (*un poco*), 3–quite something (*bastante*) to 4–a lot (*mucho*).

Attention vs Distraction

Participants from the Enhanced Captions group (n = 16) were also invited to comment upon their awareness of the enhancement in respect of attention, distraction, and memory. Several paired-samples *t*-tests were conducted to assess whether vocabulary gains were significantly different between answers.

The first of these questions asked whether they had paid more attention to the words in bold and yellow. Almost all participants (n = 14) agreed upon the fact that enhanced items had caught their attention more than those unenhanced, and these participants had significant gains at both immediate and delayed post-tests for both meaning recall ($p < .001$ and $p = .025$, respectively) and meaning recognition ($p < .001$ and $p = .003$). Participants who answered *no* (n = 2), only had significant gains at the immediate post-test for meaning recall ($p < .001$). These results suggest that paying attention to the enhanced items might have helped these participants acquire and retain the TWs better.

Secondly, concerning the level of distraction that these yellow words evoked in them, half of the participants (n = 8) stated that the enhancement distracted them not only from "the rest of the captions" but also from the documentary itself. These participants had significant gains only at the immediate post-test for meaning recall ($p = .003$) and meaning recognition ($p = .009$). The other half (n = 8) who believed that the enhancement merely caught their attention without disrupting the general comprehension had significant gains at both post-tests for both meaning recall ($p = .001$ and $p < .001$) and meaning recognition ($p < .001$ and $p < .001$). From these findings, it could be suggested that being distracted by the typographic enhancement might hinder retention of the previously acquired items.

Finally, participants answered whether they believed they had retained better those words in yellow, and their responses were predominantly in agreement. In fact, those participants that answered *yes* (n = 11) had significant gains at both the immediate and

delayed post-test, for both meaning recall ($p < .001$ and $p = .007$) and recognition ($p < .001$ and $p = .003$), whereas those that answered *no* (n = 5) did not have significant gains at the delayed post-test, only at the immediate post-test, for meaning recall ($p = .009$) and recognition ($p = .036$), which suggests that knowledge might have been further retained for those participants that indeed believed they remembered the words in yellow better.

4. Discussion

This aim of this study was to explore the effects of regular and enhanced captions as well as individual differences on incidental vocabulary acquisition –by tapping into meaning recall, meaning recognition and form recognition knowledge– through the viewing of a documentary while accounting for participants' self-reported focus of attention.

The first research question addressed the overall effects of watching the documentary on L2 vocabulary acquisition. All participants significantly gained knowledge from pre-test to either of the two post-tests for both meaning recall and meaning recognition, which suggests that viewing the captioned documentary was effective, and knowledge was significantly retained after two weeks. For meaning recall, a significant interaction between time at testing and level in which participants were enrolled (B1, B1+, B2 or C1) arose in the GLMM and thus was included in the analysis. Pairwise comparisons between level and time suggested that only for B1 and C1 participants' differences between pre-test and delayed post-test were significant. For meaning recognition, a significant interaction between vocabulary size and time was found for the slope between pre-test and immediate post-test.

For form recognition, participants' scores were significantly higher in the immediate post-test in comparison to the delayed post-

test, which suggests that participants' ability to recall having seen a particular item deteriorates with time. This is to be expected, as not encountering the items again after the viewing of the documentary does not refresh the new form-meaning connections. Interestingly enough, whereas differences between post-tests for both meaning recall and meaning recognition were non-significant, in the case of form recognition, there was a significant difference. Again, a significant main effect of vocabulary size was found for this test, but neither of LLAMA B nor D. Overall, the results of the three tests for all participants are widely consistent with previous literature, which has evidenced a positive effect of captions on L2 vocabulary learning (Gesa, 2019; Pujadas, 2019; Pujadas & Muñoz, 2019; Suárez & Gesa, 2019).

The second research question aimed at examining whether differences between types of captions occurred at the various levels of word knowledge under analysis. Results showed that there were no significant differences between caption types for the three tests at any testing time, only significant within-group differences arose. For meaning recall, on the one hand, participants who watched the documentary with EC significantly differed from their own pre-test scores in both post-tests, whereas those who viewed the documentary with RC only did from pre-test to delayed post-test. On the other hand, for meaning recognition, scores only differed significantly between testing times for the EC group. Nevertheless, differences between groups were non-significant as in other previous studies (Majuddin, 2020; Montero Perez et al.'s, 2014). As differences between immediate and delayed post-tests were non-significant for all groups, we could say that the potential learning of meaning recall and meaning recognition was not significantly lost after two weeks.

As for the form recognition test, results showed that differences between immediate post-test and delayed post-test were significant for the EC group but non-significant for the RC group. Regardless of these differences, when comparing the scores individually at each testing time, groups did not differ significantly, which is in line with

Montero Perez et al.'s (2014) findings. A significant interaction between caption type and time arose from the GLMM, but no main effects were found, in opposition to Majuddin's (2020) previous results, where the author reported that captions had significant main effects in the form recall and form recognition of multi-word expressions. In the present study, even though the EC group performed better in the first test, the RC group was able to retain more knowledge in the delayed post-test, as their scores were higher. This would suggest that the potential advantages of textual enhancement in captions is more of a short-term effect also in vocabulary gains, and not only in grammar, as was found in Pattemore and Muñoz (2022), since the EC group was not able to significantly retain the form recognition scores after two weeks.

The third research question was concerned with the extent to which vocabulary size and language learning aptitude influenced participants' scores at the different meaning and form levels. Through the different GLMMs conducted in the analysis of the study, only vocabulary size had significant main effects at all levels of knowledge for all testing times. These results are in line with most previous research, which suggests that previous vocabulary knowledge is one of the most influential factors affecting vocabulary learning (Montero Perez, 2022). As a matter of fact, it may even be suggested that the non-significant advantage in vocabulary size observed in the EC group might have helped these participants obtain an advantageous improvement, as vocabulary size did significantly interact with time at several points of the experiment. At the same time, no significant main effect arose from either one of the LLAMA tests when included as fixed factors in the statistical tests, in line with Pattemore and Muñoz (2020); not even in the learning of target word meanings, as Suárez and Gesa (2019) found.

The fourth and final research question intended to provide a quantitative and qualitative examination of learners' self-reported focus of attention. Results showed that all participants, regardless of the experimental group, reported to focus more on captions, followed

by the audio and the image, in this order. This focus on captions has also been captured through the eye-tracking methodology, with longer fixation duration associated to increased learning (Montero Perez et al., 2015). The participants also reported that their linguistic focus was mainly put on general comprehension, followed by new vocabulary, pronunciation, new expressions, and intonation. Results from the immediate post-test on comprehension confirmed that participants (all except for one) understood the documentary's essential plot, and the general gains at meaning recall and recognition for all groups goes in line with the subjects' second most appointed linguistic focus of attention, that is, new vocabulary.

Regarding the contrast between attention and distraction, which was accounted for participants in the enhanced captions group, mixed perceptions were found. First of all, almost all participants agreed upon the fact that enhanced items had caught their attention more than those unenhanced. Secondly, whereas some participants believed that the typographic enhancement did not distract them from the overall experience, some others did report having forgotten about the rest of the captions or having fixed their attention only on those words. Finally, almost all participants stated that they considered the enhancement as helpful, and most of them believed that enhanced captions was the reason behind having subsequently retained some TWs. All in all, participants were consciously aware of the typographic enhancement of certain words and, as they described in their own words, how they had noticed and, later on, acquired new vocabulary.

5. Conclusion

The current study contributes to the area of SLA through multimodal input with results from a very short exposure to a contemporary documentary during face-to-face and online classes. This study has examined the use of regular and enhanced captions

and has not found significant differences between experimental groups, although there were significant within-group differences, highlighting the relevance of out-of-classroom exposure to L2 media. Furthermore, this work has also reinforced the importance of individual differences, confirming once again the significance of vocabulary size when learning single-word items, while also studying the non-significant contribution of the LLAMA tests, which has not been widely examined in the context of caption enhancement. Finally, through the retrospective questionnaire, this study has been able to describe participants' thoughts, perceptions and ideas about the experiment in general, and about the enhancement of captions in particular.

The findings reported in this paper should be considered in the light of some limitations. Firstly, this study has not accounted for either a no-captions group or a control group who would not have viewed the documentary. Although the main aim of the study was to compare the two conditions (regular vs enhanced captions), the current findings need to be confirmed by further research including a control group. Secondly, due to the pandemic, until the final session of the experiment, half of the participants conducted the tasks online. Even though this study tried to control for cheating, and some of the websites used did not allow participants to change tabs during tests, there is an extent to which fraud cannot be fully disregarded. Another limitation of the study was that due to practical reasons the number of participants is relatively small. Further research with a larger sample size can provide stronger evidence. Next, this work could have included frequency of occurrence in the different analyses and examine its relationship to learning outcomes, previous vocabulary knowledge and language aptitude, rather than merely focusing on a frequency threshold (Uchihara et al. 2019). Further research (in progress) using the eye-tracking methodology can juxtapose students' self-reported focus of attention against more objective data to analyse the extent to which students are aware of the way in which captions attract their attention.

This study has some pedagogical implications as well. This work has demonstrated the potential advantage of multimodal input for acquiring languages. In that sense, language teachers could provide students with effective and motivating language learning experiences as some of the participants from this experiment found it "fun," "good," and "enjoyable." Altering Dr Karan Rangarajan's words from the documentary, "spread knowledge, not the virus" (Livingston, 2020, 00:17:45–00:17:47), we would like to conclude this study by asserting that we can spread culture, knowledge and languages through multimodal input, so go ahead and spread the word!

References

Feng, Y., & Webb, S. (2019). Learning Vocabulary Through Reading, Listening, and Viewing: Which Mode of Input is Most Effective? *Studies in Second Language Acquisition*, *42*(3), 499–523. https://doi.org/10.1017/S0272263119000494

Gesa, F. (2019). *L1 / L2 subtitled TV series and EFL learning: A study on vocabulary acquisition and content comprehension at different proficiency levels.* (Doctoral dissertation). University of Barcelona.

Lee, M., & Révész, A. (2020). Promoting grammatical development through captions and textual enhancement in multimodal input-based tasks. *Studies in Second Language Acquisition*, *42* (3), 625–651. https://doi.org/10.1017/S0272263120000108

Livingston, H. (Director). (2020). *Viral: The 5G Conspiracy Theory*. British Broadcasting Corporation (BBC).

Majuddin, E. (2020). *Incidental and intentional acquisition of multiword expressions from audio-visual input: The effects of typographically enhanced captions and repetition* (Doctoral dissertation). Victoria University of Wellington.

Majuddin, E., Siyanova-Chanturia, A., & Boers, F. (2021). Incidental acquisition of multiword expressions through audiovisual materials.

Studies in Second Language Acquisition, 43(5), 985–1008. https://doi.org/10.1017/S0272263121000036

Mayer, R. E. (2014). Introduction to multimedia learning. In R. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 1–24). Cambridge University Press. https://doi.org/10.1017/CBO9781139547369.002

Meara, P. M., & Miralpeix, I. (2015). *V_YesNo (v1.01)*. Cardiff: Lognostics.

Meara, P. M., & Rogers, V. E. (2019). *The LLAMA Tests v3. LLAMA_B3 (v3.00)*. Cardiff: Lognostics.

Meara, P. M., & Rogers, V. E. (2019). *The LLAMA Tests v3. LLAMA_D3 (v3.00)*. Cardiff: Lognostics.

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

Montero Perez, M., Peters, E., Clarebout, G., & Desmet, P. (2014). Effects of Captioning on Video Comprehension and Incidental Vocabulary Learning. *Language Learning & Technology*, *18*(1), 118–141.

Montero Perez, M., Peters, E., & Desmet, P. (2015). Enhancing vocabulary learning through captioned video: An eye-tracking study. *The Modern Language Journal*, *99*, 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, 720–739. https://doi.org/10.1016/j.system.2013.07.013

Muñoz, C. (2022). Audiovisual input in L2 learning. *Language, Interaction and Acquisition*. 13(1): 125-143. https://doi.org/10.1075/lia.22001.mun

Nation, P. (2001). Designing the vocabulary component of a language course. In P. Nation, *Learning Vocabulary in Another Language* (pp. 380–406). Cambridge: CUP. https://doi.org/10.1017/CBO9781139524759.013

Paivio, A. (1986). *Mental representations: A dual coding approac*h. New York: Oxford University Press.

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., & Muñoz, C. (2022). Captions and learnability factors in learning grammar from audio-visual input. *The JALT CALL Journal*, (18), pp. 83-109. http://doi.org/10.29140/jaltcall.v18n1.564

Peters, E., & Muñoz, C. (2020). Language Learning from Multimodal Input. *Studies in Second Language Acquisition*, 42, 489–497. https://doi.org/10.1017/S0272263120000212

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

Puimège, E., Montero Perez, M., & Peters, E. (2022). Promoting L2 acquisition of multiword units through textually enhanced audiovisual input: an eye-tracking study. *SECOND LANGUAGE RESEARCH*. https://doi.org/10.1177/02676583211049741

Pujadas, G. (2019*). Language learning through extensive TV viewing. A study with adolescent EFL learners*. (Doctoral dissertation). University of Barcelona.

Pujadas, G., & Muñoz, C. (2019). Extensive viewing of captioned and subtitled TV series: a study of L2 vocabulary learning by adolescents. *The 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), 1–25. https://doi.org/10.1017/S0272263120000042

Rodgers, M. P. H., & Webb, S. (2011). Narrow viewing: The vocabulary in related television programs. *TESOL Quarterly*, *45*(4), 689–717. https://doi.org/10.5054/tq.2011.268062

Rodgers, M. P. H., & Webb, S. (2019). Incidental vocabulary learning through watching television. *ITL—International Journal of Applied Linguistics*, *171*(2), 191–220. https://doi.org/10.1075/itl.18034.rod

Rogers, V., Meara, P., Barnett-Legh, T., Curry, C., & Davie, E. (2017). Examining the LLAMA aptitude tests. *Journal of the European Second Language Association*, *1*(1), 49–60. https://doi.org/10.22599/jesla.24

Schmidt, R. (1994). Deconstructing consciousness in search of useful definitions for applied linguistics. *AILA Review, 11*, 11–26.

Suárez, M. D. 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), 497–517. https://doi.org/10.1080/09571736.2019.1617768

Uchihara, T., Webb, S., & Yanagisawa, A. (2019). The Effects of Repetition on Incidental Vocabulary Learning: A Meta-Analysis of Correlational Studies. *Language Learning*, *69*(3), 559–599. https://doi.org/10.1111/lang.12343

Winke, P. M. (2013). The Effects of Input Enhancement on Grammar Learning and Comprehension. *Studies in Second Language Acquisition*, *35*, 323–352. https://doi.org/10.1017/S0272263112000903

First version received: June, 2023
Final version accepted: September, 2023