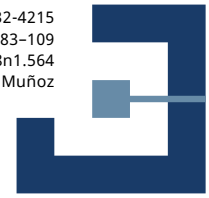
Regular Paper

The JALT CALL Journal ISSN 1832-4215
Vol. 18, No.1 Pages 83–109
https://doi.org/10.29140/jaltcall.v18n1.564



Castledown

Captions and learnability factors in learning grammar from audio-visual input

Anastasia Pattemore
anastasia.plotnikova@ub.edu
University of Barcelona, SPAIN

Carmen Muñoz
munoz@ub.edu
University of Barcelona, SPAIN

This study explores the effects of extensive audio-visual input with three captioning modes – unenhanced captions, textually enhanced captions, and no captions – on learning a variety of L2 grammatical constructions and examines the effects of three learnability factors: construction type, frequency, and recency. A total of 112 participants watched ten full-length TV series episodes over a period of five weeks. The study targeted 27 frequently occurring grammatical constructions categorized as fully-schematic, partially-filled, and fully-filled. The design included a pretest, an immediate posttest to measure the effects of recency, and a delayed posttest. The results indicated mixed effects of captioning: textually enhanced captions – a more salient condition – led to immediate learning outcomes while unenhanced captions resulted in higher long-term effects. A limit to the amount of different textually enhanced constructions presented in the input for effective learning is suggested. In general, unenhanced captions appear sufficient for successful grammar construction learning.

Keywords: audio-visual input, captions, textual enhancement, grammatical constructions

Introduction

Recent research has evidenced that TV in the target language has the potential of providing second language (L2) learners with large amounts of spoken input necessary for successful language learning (Webb, 2014, pp. 159-168). Especially, TV series – with a developing plot that encourages viewers to watch continuously – form part of an extensive viewing approach. Webb (2014) defines extensive viewing as "regular silent uninterrupted viewing of L2 television inside

and outside of the classroom". Furthermore, it has been claimed that the addition of captions – a simultaneous, on-screen written text representation of the soundtrack – benefits language learners because the three input sources (audio, visual, and caption text) complement each other and support learning from audio-visual input by distributing the information among the three channels (Vanderplank, 2016). The benefits of this multimodal input are explained by information processing theories, such as Paivio's Dual Coding Theory (1986) which outlines two independent systems – verbal and visual – which simultaneously support each other in human cognition. Partly based on this, Mayer's Cognitive Theory of Multimedia Learning (2014) assumes that people learn better when words are presented with pictures as this allows learners to make connections between word and image.

While captioning has proven itself a useful technique to foster L2 listening comprehension and vocabulary (Vanderplank, 2016), research on grammar development is still uncommon. In particular, research on the effects of sustained exposure to L2 television on grammar learning is only beginning to emerge (e.g., Pattemore & Muñoz, 2020). To address this dearth of research, we explored learning of grammatical constructions from extensive viewing of TV series under three captioning conditions: no captions, captions, and enhanced captions (a technique for raising salience). Additionally, we investigated the effects of three learnability factors of these grammatical constructions: construction type, frequency, and recency.

Background

Audio-visual input and different captioning modes

Different captioning modes and their effects on learning gains have been the subject of recent research. These captioning modes include full captions, keyword captions, textually enhanced (TE) captions, and no captions. TE captions have attracted particular attention from researchers as part of a more general interest in investigating the value of input enhancement in L2 learning (e.g., Doughty & Williams, 1998; Sharwood Smith, 1993) that is claimed to raise language salience. Textually enhanced captioned audio-visual input can be seen as a case of constructed salience, which occurs when a language feature is made more prominent (see Gass *et al.*, 2018 for definitions and types of salience in SLA).

In the context of vocabulary learning, Montero Perez *et al.* (2014) conducted an experiment with participants watching three short video clips (10'35" total) twice. All three captioning groups (full, TE, and key-word) outperformed the no captions group to some extent, and importantly, there was no significant difference between the three captioning groups. The authors concluded that salience raising by textual enhancement or key-word captioning was not more effective than traditional full captions, and suggested that the availability of captions triggered noticing of target vocabulary even without textual enhancement.

A recent study on the learning of multiword expressions (a unit longer

than a single word) through original version TV series included a comparison between no captions, captions, and TE captions (Majuddin *et al.*, 2021). The participants were exposed to a single (20 minutes) or repeated (40 minutes) viewing of one episode of a TV series. There was no group difference in the repeated viewing condition, while the results of the immediate posttest for the single viewing condition yielded a significant effect of both types of captioning over no captions, though it did not show any significant difference between unenhanced and TE captions. Furthermore, the results of the delayed posttest suggested that textually enhanced captions had a stronger effect on immediate than delayed recall. The authors argued that there was no benefit of TE captions over unenhanced captions due to the length of the multiword units – from two to five words – that could be difficult to process during the limited time that captions were on the screen. Additionally, the TE captions had to compete for attention with rapidly changing image and caption text that students were assessing while viewing the episode (compared to previous studies using TE text without audio and image).

In the context of grammar learning from textual enhancement in written texts, studies have provided inconsistent results. Some studies (Cho, 2010; Comeaux & McDonald, 2018) found textual enhancement a valuable technique, and others found no significant advantage of the TE over the unenhanced condition (Issa & Morgan-Short, 2018; Winke, 2013), suggesting that mere exposure to TE target forms without specific instructions may be insufficient to yield strong learning effects in grammar learning (Leow & Martin, 2018). This is in line with the earlier meta-analysis by Lee and Huang (2008) that reported only a small benefit of enhanced over unenhanced text (d = 0.22). Specifically, in the area of audio-visual input just four studies have focused on the effectiveness of textual enhancement on grammar learning. Lee and Révész (2018) compared the effects of TE captions (in bold) with non-enhanced captions for L2 grammatical constructions (third-person pronominal anaphora reference) learning from 27 static images with audio-recordings. The results revealed an advantage of TE over non-enhanced captions. Their later study (Lee & Révész, 2020) compared the effects of TE captions (yellow font), unenhanced captions, and no captions in directing learners' attention and learning of two L2 grammatical constructions (present perfect and past simple). The participants watched 24 short video clips (20 to 50 seconds each) in one session while their eye-movements were recorded. The results yielded significant gains in the posttest for the present perfect tense, but not for the past simple, presumably because of participants' advanced proficiency level. Both captioning modes resulted in greater learning gains, with enhanced captions being the most beneficial for learning the present perfect tense. Moreover, the eye-tracking data revealed that enhanced captions drew learners' attention to the target constructions significantly more than unenhanced captions. The authors suggested that TE captions increased visual salience of the target constructions, and subsequently students who looked at the target constructions more frequently and longer were likely to obtain higher gains in the production tasks.

Cintrón-Valentín *et al.* (2019) explored learning of L2 Spanish grammar

and vocabulary from four animated videos (specifically created for the study's vocabulary and grammar structures) under three conditions: captions with TE (bold and yellow) grammar, captions with TE (bold and yellow) vocabulary, and no-captions, but notably they did not have a purely unenhanced captions condition. A beneficial effect of TE vocabulary captions over other conditions was found for vocabulary learning. However, results were mixed for grammar. TE grammar captions and TE vocabulary captions (without highlighted grammar) showed an advantage over no captions for half of the structures. The results showed no advantage on grammar learning of the TE grammar captions condition over the TE vocabulary captions condition (without highlighted grammar). The authors argued that construction learning may depend on structure-specific saliency. Additionally, they suggested that too many grammar rules might have been presented in a single treatment video, overloading students' attention and input processing. Unfortunately, as this study lacked a pretest and included pre-teaching and pre-practice of the target grammar, it is not prudent to advocate that learning gains appeared mostly because of the audio-visual input or captions. In their follow-up study (Cintrón-Valentín & García-Amaya, 2021) the authors included a pre-test and a 'no pre-teaching' condition. They found a significant advantage of captioning for some grammar structures and discovered a significant effect of pre-teaching on grammar learning that faded over time. However, as in the previous experiment, the researchers did not have a purely unenhanced captions condition, so they could not fully compare the effects of unenhanced and enhanced captions.

To summarize, research on grammar learning from captioned audio-visual input is limited to four studies that provided only short exposure to clips, and those clips were specifically created for the interventions. They did not explore extensive exposure to L2 television, an increasingly frequent practice in many parts of the world (Muñoz, 2020; Webb, 2014). Those studies yielded mixed results concerning the benefits of different captioning modes, which might be explained by structure-specific factors (Cintrón-Valentín *et al.*, 2019). Therefore, the learnability factors of constructions may help clarify the inconsistent findings in this area.

Construction learnability factors

This study situates itself in the constructionist perspective which states that learning a language consists of the acquisition of form-meaning pairings (units) – known as constructions (Ellis & Ferreira-Junior, 2009). Ellis *et al.* (2016) state that an adult's language system is a large collection of different constructions. These units of language may differ in degrees of complexity, abstractness, transparency, and compositionality (Ellis *et al.*, 2016; Madlener, 2015). For instance, constructions carry varying levels of complexity (Pérez-Paredes, 2020) ranging from morphemes to syntactic frames; abstractness (Ellis & Cadierno, 2009) varying from concrete items (e.g. dogs) to abstractions (e.g. plurals); transparency or compositionality (Griess & Wulff, 2009) refer to whether the meanings of the separate parts of a construction represent (or do not) the whole meaning

of that construction (e.g. non-compositional *a piece of cake*, transparent *a slice of cake*).

While the Construction Grammar approach has mainly been used in L1 studies (e.g. Diessel, 2004; Goldberg, 2006), several researchers have explored L2 learning through the lens of constructions as well (e.g. De Knop, 2020; Kusyk & Sockett, 2012; Römer & Garner, 2019). Successful learning of L2 constructions depends on a number of different factors, as evidenced in several studies (Ellis & Cadierno, 2009; Ellis & Collins, 2009; Ellis *et al.*, 2016). First, constructions vary by type (Goldberg 2006), and this may play a role in their learnability. Additionally, frequency (how many times a construction appears in the input) and recency (how recently a learner has observed a construction) feature among important determinants of construction learning (Ellis & Collins, 2009).

With regard to type of construction, what counts as a construction can vary from a single morpheme (e.g. un-), to simple words, all the way up to formulaic phrases, idioms, and such complex constructions as covariational-conditional construction (e.g. 'the more, the merrier') (Goldberg, 2006). Because constructions differ in size, complexity, specificity, productivity, and interrelation, this variability creates a continuum of construction types from fixed constructions with no variation in the input, to 'slot-and-frame' or partially-filled constructions with a fixed part and a variable (schematic) slot, to schematic constructions which represent complex, highly flexible morphological or syntactic patterns (see Ellis *et al.*, 2016; Fried, 2015). There are several approaches to grouping types of constructions: Taguchi (2007) differentiates between chunks – semi-fixed grammatical patterns with one or two variable slots that carry specific functions, and unanalyzed purely formulaic expressions. Ellis (2003) distinguishes between formulae – lexical chunks that involve learning of sequences, slot-and-frame patterns – fixed grammatical frames with at least one open slot where the learners can place a variety of words, and constructions – complex chunks or high-level schemata for abstract relations (e.g. transitives, locatives, datives, passives). Using this classification, Ellis (2003) and Pérez-Paredes *et al.* (2020) suggested that second language learners learn holophrases or formulas first (e.g. *Why don't you…*) then slot-and-frame constructions (e.g. *I/you visited/went to a friend/classmate*), and finally fully abstracted formulaic chunks (e.g. *He came to the conclusion that…*). The present study adapts a similar classification of constructions by Fried (2015) and distinguishes between fully-filled (fixed multiword units with no variation in the input, e.g. *do for a living*), partially-filled (with at least one variable slot, e.g. *the Xer, the Yer*), and fully-schematic constructions (e.g. passive).

Frequency of occurrence was one of the most important factors in the learnability of grammar functors found in the meta-analysis of determinants of order of acquisition of English grammatical morphemes by Goldschneider and DeKeyser (2001). The effects of frequency have been discussed in a number of studies with a constructionist perspective (e.g. Ellis & Ferreria-Junior, 2009), and recently also in research into audio-visual input.

Specifically, Muñoz *et al.* (2021) explored learning of vocabulary and abstract constructions from extensive audio-visual exposure to TV series. Frequency of

occurrence of vocabulary was positively correlated with learning outcomes, supporting previous evidence found of frequency being a potential predictor of vocabulary learning from L2 audio-visual input (e.g. Peters, 2019; Peters & Webb, 2018). However, the vocabulary correlations were smaller than in most previous vocabulary studies. The authors suggested the frequency effects may have been attenuated by the combination of on-screen text and visual images (as observed in the meta-analysis by Uchihara *et al.*, 2019). Likewise, Pellicer-Sánchez (2016), in her study on effects of frequency on collocation learning while reading, did not find a significant effect of collocation frequency and suggested that the effect of frequency might be influenced by other factors (such as spacing of exposure, see Uchihara *et al.*, 2019). Regarding the learning of constructions, the results showed that the association between constructions' frequency of occurrence in the input and learning outcomes was much higher when the audio-visual input was presented without captions than with captions. That is, frequency effects were significant in the more challenging condition of the study.

Recency of occurrence is a relatively unexplored factor in both construction grammar and audio-visual input research. It can be defined as the time since past occurrence of a stimulus (Robinson *et al.*, 2012) and it is one of the key factors in activating memory schema (Ellis & Collins, 2009). According to Ellis (2006), both cognition and memory are sensitive to recency: "the probability of recalling an item, like the speed of its processing or recognition, is predicted by time since past occurrence" (p.5). Therefore, a learner's memory would be stronger of a construction more recently presented in the input and, consequently, it would be accessed more fluently (Ellis, 2012).

Aim and research question

This study aims to examine the potential benefits of different types of captioning on grammar learning and the effects of three factors that may be determinants of construction learnability: construction type, frequency, and recency. To the best of our knowledge, this is the first study to address L2 grammar learnability factors from extensive exposure (five weeks of regular uninterrupted viewing) to full-length TV series episodes with captions, enhanced captions, and no captions conditions. The following two-part research question is addressed in this study:

- Research Question 1a: To what extent does L2 grammatical construction learning from audio-visual input depend on caption support, construction type, frequency, and recency?
- Research Question 1b: In what ways do these factors interact in L2 grammatical construction learning from audio-visual input?

In this study the construction type is operationalised as three groups, based on Fried (2015), as described above: fully-filled, partially-filled, and fully-schematic. Frequency is measured by number of times target constructions appear in the chosen TV series episodes. Caption support is operationalised as three

groups varying in the type of on-screen text: no captions, captions, and textually enhanced captions (raised salience mode). Finally, recency of occurrence is operationalised as test recency: the comparison between learning gains on the immediate and delayed posttests.

Methodology

Participants

The initial pool of participants was 149 Catalan/Spanish bilingual undergraduate students from a university in Spain. They were attending obligatory English classes as a part of their Audiovisual Communication degree, and they had not been streamed by L2 level. Students were given class credits for watching the episodes and completing the tests. The participants (n = 37) who did not watch all the episodes or did not complete the required tests were excluded from the analysis, leaving 112 participants.

All groups had the same language teacher and were not informed about the nature of the experiment beforehand. The content of the class was vocabulary based and related to students' major (advertising, cinema, marketing), the curriculum did not cover grammar practice, and the teacher was explicitly asked not to provide any instruction of the target grammar constructions. The participants' English proficiency varied from B1 to C1 (with a mean of B2, see Table 1) according to the Common European Framework of Reference for Languages (CEFR) (Council of Europe, 2001).

Four intact classes were randomly assigned to one of four conditions: Captions (n = 32), Captions (-) (n = 30), No Captions (n = 22), and TE Captions (n = 28). The Captions (-) group did not complete the immediate posttests; it was included to account for a possible testing effect from the immediate posttest (see below).

Target constructions

The target constructions (TCs) addressed in this study were selected from the script of the ten consecutive episodes of the first season of the comedy TV series *The Good Place* (Schur, 2016). A total of 27 constructions were targeted based on their frequency of occurrence: only constructions that appeared at least three times in the ten episodes (227 minutes) were included in the study. The class teacher was also consulted to determine whether the selection of constructions was appropriate given the wide range of proficiency existing in the groups. Given that constructions are learnt through the specific input which learners are exposed to (Goldberg & Casenhier, 2008), the TCs were allocated to three groups depending on the specific language exemplars present in the input material: fully-schematic (10), partially-filled (11), and fully-filled (6). See Appendix A for a list of TCs, their categorisation, question type in the tests, and frequency of occurrence.

The application "Subtitle Edit" (Version 3.5.10; Olsson, 2019) was used to

enhance TCs for the TE Captions group. The constructions were highlighted in yellow and bold. Each episode featured 14 to 40 uses of TCs, with 7 to 16 different TCs per episode, representing about 4.5% of highlighted text in each episode. Only one construction was highlighted at a time; if there were two or more TCs appearing on the screen at the same time, then the construction with lower frequency was highlighted to avoid students splitting their attention between multiple enhanced constructions presented on the screen (Ayres & Sweller, 2014).

Testing materials

A pen-and-paper version of the Oxford Placement Test (OPT) (Allan, 2004) was used to measure participants' proficiency. This test consists of two sections: listening comprehension and grammar, with a total score of 200 which can be converted to CEFR proficiency levels.

The testing materials for the pre- and delayed posttests consisted of 54 test items (two test items per TC) and included productive grammar exercises, such as sentence transformation, fill-the-gap, and complete the gap with a correct form of a given word (see Appendix B). These types of exercises were chosen due to students' familiarity with the format (from language coursebooks). To pilot and validate the test items, two native English speaking EFL teachers completed the test to see whether the items would elicit use of the TCs. If the test items elicited non-target constructions, then the items were changed and tested again. Later, all test items were piloted with a comparable group of participants (n = 15) and further alterations were made, for example the wording of the instructions was clarified, and unfamiliar vocabulary was changed to known synonyms. It is important to note that the test items did not present the TCs to students; it was the students who had to analyse the prompt and produce the TCs themselves. Both pre- and delayed posttests contained the same test items in a randomized order. Cronbach's alpha showed that both the pretest (α = 0.879) and the delayed posttest (α = 0.835) reached an acceptable level of internal consistency.

An immediate posttest was administered in the second class of each week. The aim of this test was to contrast the immediate construction learning with cumulative learning in the delayed posttest as one measure of recency. The immediate posttest included test items on five to seven TCs that appeared with the highest frequency in the preceding two episodes in that week. Those tests had the same format as pre-/delayed post- tests (productive grammar exercises) but had different test items to avoid practice for the delayed posttest. In total, the five partial immediate posttests consisted of 54 test items with two items per construction, every construction was tested once throughout the partial immediate posttests (see Appendix C). The Cronbach's alpha for the immediate posttest reached an acceptable reliability level (α = 0.795).

Procedure

The procedure is summarized in Figure 1. The intervention took place twice a week (90 minutes per class) over a period of eight consecutive weeks. The first two weeks students completed the OPT (60 minutes) and the pretest (40 minutes). The following five weeks the participants watched two full-length episodes of the TV series on two different days. Every week followed the same protocol: for practical reasons, the first day students watched one episode and completed comprehension questions (10 minutes, see Appendix D for an example); the second day students watched the next episode and completed comprehension questions (10 minutes) together with the immediate posttest (10 minutes). The Captions (-) group followed the same protocol but did not complete the immediate posttest and continued with the class instead. This allowed us to control a potential testing effect of the immediate posttest on the delayed posttest. Students completed the delayed posttest in the last week of the intervention, five days after watching the last episode.

One point was assigned for a correct answer per question in the pretest, immediate posttest, and delayed posttest. An answer was considered correct when it included all parts of a construction in the correct form, but the students were not penalized for spelling mistakes.

The Captions and the Captions (-) groups watched the TV series with English captions, the TE Captions group watched the episodes with English captions with the TCs in bold and yellow, and the No Captions group watched the episodes without captions.

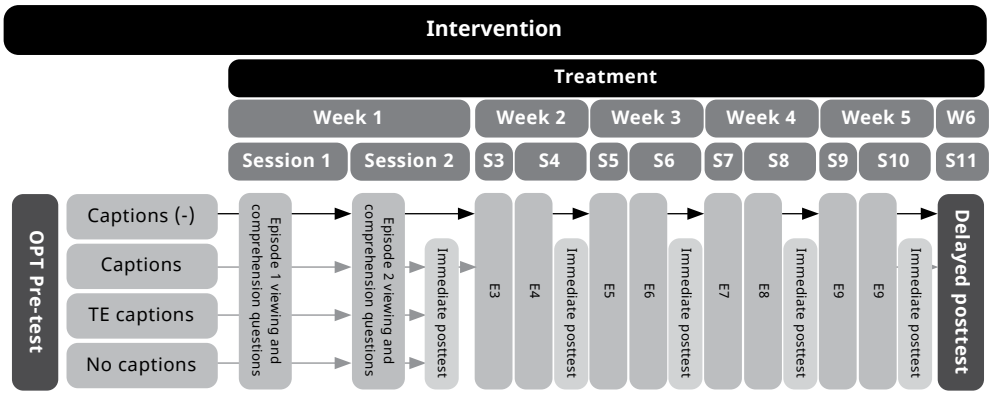


Figure 1. Pedagogical intervention

Results

Preliminary analysis

The initial exploration of the data showed that there were no significant differences between the four groups in terms of overall proficiency (F (3,108) = 1.533, p = .210), listening part of proficiency test (F (3, 108) = .2.366, p = .075), grammar part of proficiency test (F (3,108) = .557, p = .645) (see Table 1), or pretest scores (F (3,108) = .311, p = .817) (see Table 2).

As mentioned above, the reason the Captions (-) group did not take the

immediate posttest was to control for a possible testing or practice effect. To investigate this, an independent samples t-test was run with the delayed posttest scores of the two Captions groups: Captions and Captions (-). The results showed that there was no significant difference between them (t (60) = .213, p = .832). Thus, a testing effect resulting from the immediate posttest itself was not observed. The Captions (-) group was not included in the main analyses as these included the immediate posttest scores.

Table 1. Proficiency scores

		General proficiency (max: 200)	Listening (max: 100)	Grammar (max: 100)
Group	**N**	**Mean (SD)**	**Mean (SD)**	**Mean (SD)**
Captions	32	148.00 (15.67)	75.81 (7.23)	72.18 (10.49)
TE Captions	28	139.89 (12.96)	70.96 (6.16)	68.92 (9.66)
No Captions	22	142.54 (17.17)	74.04 (8.58)	71.50 (11.59)
Captions (-)	30	144.93 (13.68)	73.43 (6.50)	71.50 (9.73)
All	112	144.66 (14.94)	73.61 (7.21)	71.05 (10.26)

Note. Captions (-) = captions with no immediate posttest

Table 2. Pretest, immediate posttest, and delayed posttest scores

		Pretest (max: 54*)		Immediate posttest (max.: 54)			Delayed posttest (max.: 54)			
Group	**N**	**Mean (SD)**	**95% CI**	**Mean (SD)**	**95% CI**	***d*¹**	**Mean (SD)**	**95% CI**	***d*²**	***d*³**
Captions	32	23.31 (9.80)	[19.77, 26.84]	33.81 (9.44)	[30.40, 37.21]	1.11	37.68 (9.69)	[34.19, 41.18]	1.48	0.40
TE Captions	28	24.39 (8.98)	[20.90, 27.87]	37.39 (6.83)	[34.74, 40.04]	1.90	35.00 (8.65)	[31.64, 38.35]	1.22	0.27
No Captions	22	25.86 (9.15)	[21.80, 29.92]	36.22 (8.00)	[32.67, 39.77]	1.29	35.36 (9.41)	[31.18, 39.53]	1.00	0.09
Captions (-)	30	24.80 (10.79)	[20.76, 28.23]	–	–	–	37.16 (9.51)	[33.61, 40.72]	1.29	–
All	112	24.48 (9.67)	[22.67, 26.29]	35.68 (8.29)**	[33.85, 37.50]	1.35	36.41 (9.28)	[34.68, 38.15]	1.28	0.07

* two test items per 27 constructions

**n = 82

[1] Cohen's *d* for difference between the pretest and immediate posttest scores

[2] Cohen's *d* for difference between the pretest and delayed posttest scores

[3] Cohen's *d* for difference between the immediate and delayed posttest scores

Research question 1a: Factors explaining construction learning

The first sub-question focused on the effects of captions, construction type, frequency, and test recency, in construction learning from audio-visual input. The descriptive statistics for test scores are presented in Table 2 (the *n* for immediate posttest is smaller because Captions (-) did not take this test).

A series of LMMs were fitted in R version 3.6.3 (R Core Team, 2020) using the lmer() function from the *lme4* package (Bates *et al.*, 2015) and using restricted maximum likelihood to perform an LMM analysis of the relationship between test outcomes and learnability factors. The LMMs were fitted with the pretest, immediate posttest, and delayed posttest raw scores (continuous score at item level) divided by maximum possible score in the test as a dependent variable. We had to add the maximum possible test score in the analysis because this study's groups varied in number of participants. To answer the first subquestion, fixed effects included captioning mode (Captions, No Captions, TE Captions), construction type (fully-schematic, partially-filled, fully-filled), frequency of occurrence, and time (pretest, immediate posttest, delayed posttest). Each construction was included as a random subjects effect in the model. The R scripts and packages used are reported in Appendix E.

The first model carried out was an unconditional means model to see whether LMMs were a suitable type of analysis for this dataset. The construction variance component was significant ($p < .001$) in this null model and therefore the multilevel modeling was concluded to be appropriate for this data analysis.

The second model explored the relationship between fixed effects and the dependent variable. This model revealed significant fixed effects of construction type (χ^2 (2) = 11.828, p = 0.002), and time (χ^2 (2) = 245.959, $p < .001$). The estimated marginal means are reported in Table 3.

Table 3. Estimated marginal means of fixed effects

Fixed effect	Levels	Estimated marginal mean (SE)	df	95% CI
Construction type	Fully-schematic	0.659 (0.061)	33.1	[0.535, 0.784]
	Partially-filled	0.705 (0.060)	33.1	[0.582, 0.829]
	Fully-filled	0.380 (0.083)	33.1	[0.210, 0.551]
Time	Pretest	0.466 (0.038)	36.2	[0.387, 0.545]
	Immediate posttest	0.639 (0.038)	36.2	[0.560, 0.718]
	Delayed posttest	0.640 (0.038)	36.2	[0.561, 0.719]

The results suggest that there was no significant difference between learning fully-schematic and partially-filled constructions by all participants (estimate = −0.046, SE = 0.084, p = .848) but that for both fully-schematic (estimate = 0.278, SE = 0.106, p = .034) and partially filled (estimate = 0.325, SE = 0.106, p = .013) construction types, participants correctly answered between a quarter and a third more of the available questions than for the fully-filled

constructions, suggesting that fully-filled constructions were learnt the least from this intervention.

Regarding the time difference, all participants improved their knowledge of constructions between the pretest and immediate posttest (estimate = −0.172, SE = 0.012, p <.001) and between the pretest and the delayed posttest (estimate = −0.174, SE = 0.012, p <.001), scoring on average 17% higher in both posttests. There was no significant difference between the immediate and delayed posttests' scores in this model (estimate = −0.001, SE = 0.012, p = .992).

No significant fixed effect of group was observed (χ^2 (2) = 2.331, p = 0.311), suggesting that learning outcome could not be solely explained by viewing condition. There was also no significant effect of frequency on learning (χ^2 (1) = 0.117, p = 0.732). The conditional (R^2c = *.881)* and marginal (R^2m = *.358)* R^2 demonstrated that the whole model and the fixed effects respectively explained a large amount of the variance in the dependent variable.

Research question 1b: interaction between construction learnability factors

The second sub-question focused on the interactions between the different construction learnability factors, including interactions between group and construction type, group and frequency, and group and recency (test time).

The group by construction interaction did not reach significance (χ^2 (2) = 8.226, p = .083), suggesting that there was no group difference in learning different construction types (see Figure 2). The group by frequency did not have a significant effect on learning either (χ^2 (2) = 4.080, p = .130).

As for the testing time and group differences, the interaction between group and time was found to be significant (χ^2 (4) = 23.379, p < .001). The results of this interaction are presented in Figure 3. As in the first model, the conditional (R^2c = *.898)* and marginal (R^2m= *.335)* R^2 showed that the model accounted for a large amount of variance in the dependent variable.

Table 4. Estimated marginal means of captioning mode per time of testing

Time	Group	Estimated marginal mean (SE)	df	95% CI
Pretest	Captions	0.444 (0.041)	41.7	[0.361, 0.527]
	No Captions	0.475 (0.041)	41.7	[0.392, 0.558]
	TE Captions	0.476 (0.041)	41.7	[0.392, 0.559]
Immediate posttest	Captions	0.607 (0.041)	41.7	[0.523, 0.690]
	No Captions	0.627 (0.041)	41.7	[0.544, 0.711]
	TE Captions	0.679 (0.041)	41.7	[0.595, 0.762]
Delayed posttest	Captions	0.676 (0.041)	41.7	[0.593, 0.759]
	No Captions	0.621 (0.041)	41.7	[0.537, 0.704]
	TE Captions	0.620 (0.041)	41.7	[0.537, 0.704]

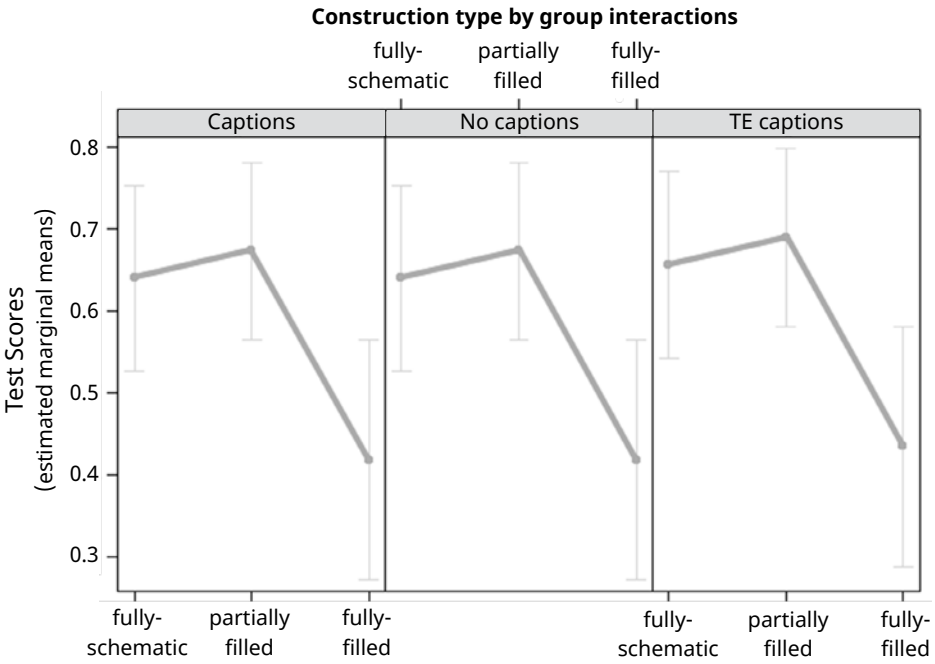


Figure 2. Test scores (divided by maximum possible score) by captioning mode group and construction type

Regarding the pairwise comparison of the effects of captioning modes on immediate posttest scores, the model revealed 7% higher scores for the TE Captions group than for the Captions group (estimate = 0.072, SE = 0.021, p = .002), and a significant difference between the TE Captions group and No Captions group with the TE Captions group scoring 5% higher (estimate = 0.051, SE = 0.021, p = .046). No significant difference was observed between the Captions and No Captions groups (estimate = -0.020, SE= 0.021, p = .596). With respect to the delayed posttest scores, the Captions group outperformed both the TE Captions (estimate = 0.055, SE = .0.21, p = .027) and No Captions groups (estimate = .055, SE = .021, p = .028) by about 5%, learning around 1 or 2 more new constructions. There was no significant difference between the TE Captions and No Captions delayed posttest scores (estimate = 0.000, SE = 0.021, p = 1.000). These results are summarized in Table 5.

Table 5. Coefficients of pairwise contrasts of group by time interaction, group comparison

Time	Group contrast	Estimate	SE	Df	t	*p*
Immediate posttest	Captions – No Captions	−0.020	0.021	231	−0.970	0.596
	Captions – TE Captions	−0.072	0.021	231	−3.360	0.002
	No Captions – TE Captions	−0.051	0.021	231	−2.390	0.046
Delayed posttest	Captions – No Captions	0.055	0.21	231	2.580	0.028
	Captions – TE Captions	0.055	0.021	231	2.588	0.027
	No Captions – TE Captions	0.000	0.021	231	0.008	1.000

The results concerning test recency per group (see Table 6 and Figure 3) – the comparison between the short-term and long-term learning – showed that the Captions group had 7% greater scores in the delayed than in the immediate posttest (estimate = 0.069, SE = 0.021, p = .003) and conversely, the TE Captions group showed 6% higher scores in the immediate than in the delayed posttest (estimate = 0.058, SE = 0.021, p = .017). Finally, the No Captions group did not have significantly different scores in the posttests (estimate = 0.006, SE = 0.021, p = .945).

Table 6. Coefficients of pairwise comparison of group by time interaction, time comparison

Group	Contrast	Estimate	SE	Df	t	*p*
Captions	Pretest – Immediate posttest	−0.162	0.021	231	−7.678	<. 001
	Pretest – Delayed posttest	−0.232	0.021	231	−10.957	<. 001
	Immediate posttest – Delayed posttest	−0.069	0.021	231	−3.279	0.003
No Captions	Pretest – Immediate posttest	−0.152	0.021	231	−7.193	< .001
	Pretest – Delayed posttest	−0.145	0.021	231	−6.876	< .001
	Immediate posttest – Delayed posttest	0.006	0.021	231	0.318	0.945
TE Captions	Pretest – Immediate posttest	−0.203	0.021	231	−9.587	< .001
	Pretest – Delayed posttest	−0.144	0.021	231	−6.839	< .001
	Immediate posttest – Delayed posttest	0.058	0.021	231	2.748	0.017

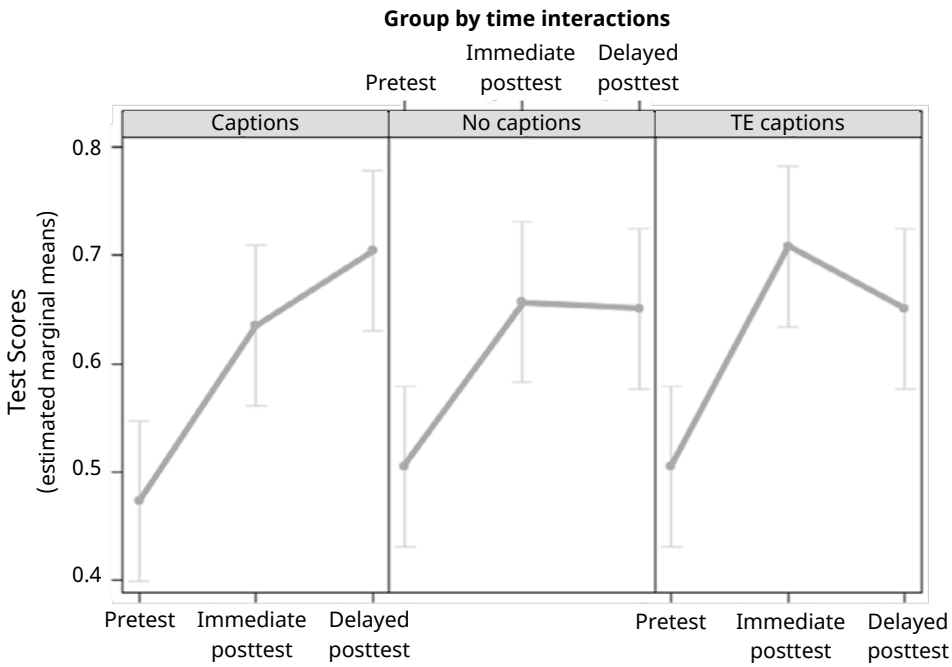


Figure 3. Test scores (divided by maximum possible score) by captioning mode and time of testing

To summarize the results from the two sub-questions, in the immediate posttest the TE Captions group significantly outperformed both Captions and No Captions groups, and there was no difference between the Captions and No Captions groups' scores. Contrariwise, the Captions group demonstrated the greatest scores in the delayed posttest, and the significant difference disappeared between the TE Captions and No Captions groups. Regarding construction type, the partially-filled and fully-schematic constructions were learnt significantly better than fully-filled constructions. Frequency of occurrence did not have a significant effect on learning of either of the groups. Regarding test recency, the Captions group had significantly higher scores at the moment of the delayed than of the immediate posttest, while TE Captions had higher scores in the immediate posttest. Finally, the No Captions group's scores did not differ significantly between the tests.

Discussion

The present study was designed to investigate to what extent extensive audio-visual input could support L2 construction learning. In particular, this study is the first to investigates the effects on such learning of captioning mode, construction type, frequency, and recency. The analysis of the overall effect of audio-visual input on L2 construction learning showed that all groups, regardless of the captioning condition, significantly improved their knowledge of the TCs. This goes in line with theories supporting learning from multimodal input (e.g. Paivio, 1986; Mayer, 2014) that explain that audio and image sources concurrently support audio-visual input processing, resulting in better learning outcomes. Additionally, the results are in accordance with the general benefit of captioned audio-visual input on language learning suggested by various studies (see Vanderplank, 2016), and the specific benefit on grammar learning shown in Lee and Révész (2018, 2020).

Our general research question was divided into two specific ones. The first sub-question explored whether captioning mode, construction type, frequency of construction occurrence, and recency affected learning of TCs. The second sub-question focused on the ways in which these factors interact in L2 construction learning from audio-visual input.

Factors explaining construction learning

Construction type. Regarding construction type, we distinguished between fully-schematic, partially-filled, and fully-filled constructions (Fried, 2015). Our results revealed that not all constructions were learnt to the same degree; partially-filled and fully-schematic constructions were learnt significantly better than fully-filled constructions in both immediate and delayed posttests. This might suggest that partially-filled and fully-schematic constructions – less constrained (and thus easier to use) and more productive than the fully-filled constructions (which can often only be used in a single manner) – are easier to learn from audio-visual input. Ellis (2003) and Pérez-Paredes *et al.* (2020)

suggested that the acquisition of L2 constructions followed a specific order: from formulae, to slot-and-frame constructions, and then to fully abstracted formulaic chunks. Although we used a different classification in this study, our categories have common features to the ones mentioned above. However, our results do not support this order of constructions acquisition. The fully-filled constructions, fixed lexical chunks or formulae, were not acquired first or better than either of the construction types that are suggested to be acquired at later stages. Additionally, there was no difference between the fully-schematic and partially-filled construction learning. Considering our participants' high intermediate level of English it might be possible that they were equally ready to acquire both partially-filled and fully-schematic constructions to the same degree, while the fully-filled constructions in this specific audio-visual input may have been not salient, relevant or frequent enough. However, much more research is needed to explore whether an order of L2 construction learning from audio-visual input can be established.

Frequency. The results from the study also showed that construction frequency did not have a significant effect on learning outcomes. This clashes with previous research demonstrating a significant association between frequency of occurrence and grammatical construction learning from audio-visual input for the no captioned group but not for the captioned group (Muñoz *et al.*, 2021). However, our results are in line with Pellicer-Sánchez's (2017) claim that frequency effects might be overpowered by other factors. For instance, this difference in results may lie in a greater variety of proficiency levels in the study by Muñoz *et al.* (2021) where elementary proficiency students were included in the analysis. We may hypothesise that lower proficiency students exposed to non-captioned video benefit from the external support of frequency more than higher level students. This would explain the smaller effect of frequency in the present study with more advanced learners.

Captioning mode and cumulative learning . As regards the effects of captioning mode on learning from extensive exposure to ten full-length episodes of a TV series, as seen above, the delayed posttest showed that the Captions group outperformed both the TE Captions and No Captions groups. This supports the previously demonstrated benefit of captioned over non-captioned audio-visual input for L2 grammar learning (Lee & Révész, 2020). In the present study, the students who watched the TV series with unenhanced captions benefited from the full intervention the most, as shown by the delayed posttest scores, in contrast to the studies by Lee and Révész (2018, 2020) where enhanced captions led to higher gains than unenhanced captions. Additionally, the TE Captions group did not significantly attain more than the No Captions group which runs counter to the findings in the study by Lee and Révész (2020), but partially confirms the mixed results in the studies by Cintrón-Valentín *et al.* (2019) and Cintrón-Valentín and García-Amaya (2021). The finding that the TE Captions – a more salient condition – did not outgain the rest of the groups at the end of the intervention partly harmonizes with the study of Montero Perez and

colleagues (2014) on vocabulary learning from audio-visual input. In that study more salient conditions – enhanced captions and keyword captions – did not lead to higher learning gains compared to unenhanced captions. The authors suggested that the captions themselves already increase the salience of the target items. Our results on grammatical constructions add to their mixed findings, and it may be suggested that additional highlighting of the TCs might have been unnecessary and not attracted enough extra attention to promote better learning and exceed the rest of the groups.

Conversely, a thought-provoking explanation for our results may lie in differences in the characteristics of the grammar experiments. Our study looked at prolonged exposure to media in the target language, while previous grammar studies exposed their participants to audio-visual materials specifically created for the interventions that took relatively short periods of time. It might be that those studies (Lee & Révész, 2018; 2020; Cintrón-Valentín *et al.*, 2019; Cintrón-Valentín & García-Amaya, 2021) captured the immediate benefits of enhanced captions, having their posttests immediately after viewing, while the present study captured longer-term benefits of captions as well. It could be suggested by our results that salience raising by textual enhancement has more immediate than cumulative or long-term effects, which would also be supported by the higher scores of the TE group in the immediate than in the delayed posttest. Another explanation may lie in the type of audio-visual materials used in different studies. Our results align with Majuddin *et al.* (2021) where participants watched a 20-minute original version episode of a TV series and there was no significant difference between the unenhanced captions and TE captions at the end of the intervention. The authors elaborated that it could be a result of the fast-paced and dynamic nature of authentic TV series when the enhanced captions only appear on the screen briefly and have to compete with Hollywood stars and special effects, compared to static images or animated videos created specifically for classroom purposes.

A second aspect that differentiates the present study from previous ones is the number of grammatical constructions involved in the learning process. While in the studies by Lee and Révész (2018, 2020), Cintrón-Valentín *et al.* (2019), and Cintrón-Valentín and García-Amaya (2021) the focus of the clips was on either one construction at a time or a contrast between two structures, the present study focused on 27 different constructions which were presented simultaneously throughout the episodes (from 7 to 16 different TCs per episode). Likewise, Majuddin *et al.* (2021) also targeted multiple multiword units (18) in a single episode and there was no benefit of TE captions over unenhanced captions. Cintrón-Valentín and colleagues (2019) suggested that a contrast between the grammar structures along with textual enhancement in a single treatment video might overload students' input processing and attention and therefore TE captions may be more effective when directed to one grammatical form at a time. The results from the present study seem to lend support to this claim and indicate the effect of attention limitations at work when a number of textually enhanced constructions are presented simultaneously in the input.

Interaction between captioning mode and construction learnability factors. Finally, the second sub-question examined the effect of the interaction between construction type with group, frequency of construction occurrence with group, and test recency with group on the learning of TCs. The interaction between the construction type and captioning mode yielded no significant results. It seems that in our study learning of different types of constructions did not depend on the captioning mode. As mentioned above, the results regarding construction type and audio-visual input are initial and more research is needed to unveil whether certain construction types are learnt better under various viewing conditions. Similarly, there was no significant difference between the frequency of occurrence and captioning mode. As discussed earlier, this lack of association might be a result of other factors such as proficiency playing a more crucial role in the learning of TCs.

As for the test recency with group interaction, we compared scores from the immediate and delayed posttests of the three groups in this study. Interestingly, the results showed that the three groups went in different directions. The Captions group demonstrated significantly higher scores in the delayed posttest than in the immediate posttest, suggesting that the long-term benefit from exposure to captions, i.e., the cumulative amount of encounters with the captioned TCs, may be higher than the immediate benefit, at least for the time periods in this study. In contrast, the TE Captions group achieved significantly higher results in the immediate posttest than in the delayed posttest, and it had higher scores than the Captions group in the immediate posttest; that is, TE captions appeared more valuable in the short term. Finally, the No Captions group neither significantly improved nor worsened between the tests. Interestingly, our TE Captions group had a significant advantage over the No Captions group in the immediate posttest, but did not have higher scores in the delayed posttest.

This is in line with Ellis's (2015, p. 171) suggestion that input enhancement does not always have a positive effect on learning; especially in the case of overenhancement it could have a damaging effect. Therefore, one explanation for the finding of only a short-term benefit of TE Captions could lie in the challenge imposed by the large number of TCs in the input (the TCs appeared from 14 to 40 times in a single episode) leading to overenhancement. Possibly even 4.5% of highlighted text was excessive and this constrained students' processing of the target structures, resulting in lower learning gains (Han *et al.*, 2008). Although we presented only one TC at a time to avoid split attention (Ayres & Sweller, 2014), the number of constructions highlighted and targeted in a single episode or intervention may simply have been too large. In this vein, attentional processing of enhanced constructions should be explored with eye-tracking measures to see at what point enhanced captions stop receiving students' focused attention. For instance, the length of students' fixations on the target items may reveal whether the intervention effect is diminishing over time and if the constructions appearing at the beginning of an episode receive more attention than those presented towards the end. This could help us shed light on the relative amount of textual enhancement for captions that is optimal in pedagogic materials.

Limitations and further research

The study is not without limitations. The first lies in the testing materials, as can be seen in Appendix A, the constructions were not split evenly between the different types of exercises because we were using authentic audio-visual input and could not control for an even number of constructions of the various types. Another limitation of the study may be seen in the large number of TCs that, while allowing for a more thorough exploration of the learnability of the different types of constructions, may have made the learning task very challenging. It could be suggested that future studies could develop a condition where TCs are enhanced only the first time they appear in the episode, thus decreasing frequency of textual enhancement and examining whether this would promote learning. Finally, the use of the immediate posttest added some practice that might have enhanced construction learning, although such practice did not provide any feedback, happened only once per TC, and could not be observed through the statistical analysis. Conversely, increasing the number of immediate posttests by having one after each episode might have allowed us to measure effects of recency more precisely.

Conclusion

This study extends the benefits of audio-visual input for grammar learning to an extensive intervention in a classroom situation. The results of this study make original contributions and provide evidence of the ways in which learning outcomes are influenced by captioning mode, construction type, frequency, and recency of exposure. In general, ordinary captions led to higher cumulative learning outcomes from extensive exposure to the L2 audio-visual input, while TE captions had an immediate effect on L2 construction learning that faded over time. Several insights were also obtained concerning construction learnability factors. Firstly, construction type was shown to be a crucial factor with fully filled constructions being learnt to a lesser extent by all groups. Secondly, the frequency of construction occurrence did not seem to have an effect on intermediate and advanced students' learning. The third insight lies in the diverse effects of recency on different captioning modes.

The study also has implications for language teaching and learning. First, although the participants significantly improved their constructions knowledge with a medium to large effect size, the actual raw number of learnt constructions was only between 5 to 7 constructions depending on the group (see Table 2). This was not unexpected from an intervention that led to incidental learning. Teachers may use audio-visual material in the classroom with a focus on form (see for example, Pujadas & Muñoz, 2019) as well as motivate learners to view audio-visual input extensively outside the classroom to increase the amount and quality of L2 input. Second, as noted above, where enhanced captions are used, only a limited number of items should be targeted to avoid attention limitations imposed by the simultaneous presentation of various enhanced constructions. However, if the findings of this study are corroborated by further research, teachers may not need to manipulate captions. Studies seem

to be showing that long-term grammar learning outcomes can be achieved through the use of audio-visual input without caption manipulation consuming too much of a teacher's precious time. It appears that unenhanced captions (which are already available on most media platforms) are advisable to be used both in and outside of the classroom.

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. We are grateful to María del Mar Suárez for her help with the data collection, and to Matthew Pattemore for proof-reading the article.

References

Allan, D. (2004). Oxford Placement Test 1. Oxford University Press.

Ayres, P., & Sweller, J. (2014). The split-attention principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (p. 206–226). Cambridge: Cambridge University Press.

Bates D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software,* 67, 1–48.

Buchsbaum, B.R. (2016). Working memory and Language. In G. Hickok, & S.L. Small (Eds.), *Neurobiology of Language* (pp. 863–875). Academic Press.

Barton, K. (2020). *MuMIn: Multi-Model Inference// R package version 1.43.17.* https://cran.r-project.org/web/packages/MuMIn/

Council of Europe (2001). Common European Framework of Reference for Languages: Learning, teaching, assessment. Cambridge: Cambridge University Press.

Cho, M.Y. (2010). The effects of input enhancement and written recall on noticing and acquisition. *Innovations in Language Learning and Teaching,* 4, 71–87.

Cintrón-Valentín, M., & García-Amaya, L. (2021). Investigating textual enhancement and captions in L2 grammar and vocabulary. *Studies in Second Language Acquisition*. 1–26. https://doi.org/10.1017/S0272263120000492.

Cintrón-Valentín, M., García-Amaya, L., & Ellis, N. C. (2019). Captioning and grammar learning in the L2 Spanish classroom. *The Language Learning Journal,* 47(4), 1–21.

Comeaux, I., & McDonald, J.L. (2018). Determining the effectiveness of visual input enhancement across multiple linguistic cues. *Language Learning*, 68(1), 5–45.

De Knop, S. (2020). The embodied teaching of complex verbal constructions with German placement verbs and spatial prepositions. *Review of Cognitive Linguistics*, 18(1), 131–161.

Diessel, H. (2004). *The acquisition of complex sentences*. Cambridge: Cambridge University Press.

Doughty, C., & Williams, J. (Eds.) (1998). *Focus on form in classroom second language acquisition*. Cambridge: Cambridge University Press.

Ellis, N.C. (2003). Constructions, chunking, and connectionism: The emergence of second language structure. In C. Doughty & M. H. Long (Eds.), *The Handbook of second language acquisition* (pp. 63–103). Blackwell Publishing.

Ellis, N.C. (2006). Language acquisition as rational contingency learning. *Applied Linguistics*, 27(1), 1–24.

Ellis, N.C. (2012). Frequency-based accounts of second language acquisition. In M.S. Gass, & A. Mackey (Eds.), *The Routledge Handbook of Second Language Acquisition* (pp. 193–210). Routledge.

Ellis, N.C., & Cadierno, T. (2009). Constructing a Second Language: Introduction to the Special Section. *Annual Review of Cognitive Linguistics*, 7, 111–139.

Ellis, N.C., & Collins, L. (2009). Input and second language acquisition: The roles of frequency, form, and function. Introduction to the special issue. *The Modern Language Journal*, 93(3), 329–335.

Ellis, N.C., & Ferreira-Junior, F. (2009). Construction learning as a function of frequency, frequency distribution, and function. *The Modern Language Journal*, 93(3), 370–385.

Ellis, N.C., Römer, U., & O'Donnel, M. (2016). Constructions and usage-based approaches to language acquisition. *Language Learning,* 66(Supplement 1), 23–44.

Ellis, R. (2015). *Understanding second language acquisition.* Oxford: Oxford University Press.

Fox, J., Weisberg, S., Price, B., Adler, D., Bates, D., Baud-Bovy, G., Bolker, B., Ellison, S., Firth, D., Friendly, M., Gorjanc, G., Graves, S., Heiberger, R., Krivitsky, P., Laboissiere, R., Maechler, M., Monette, G., Murdoch, D., Nilsson, H., Ogle, D., Ripley, B., Venables, W., Walker, S., Winsemius, D., & Zeileis, A. (2020). *Car: Companion to applied regression // R package version 3.0-10.* Retrieved from https://cran.r-project.org/web/packages/car/

Fried, M. (2015). Construction grammar. In T. Kiss, & A. Alexiadou (Eds.), *Syntax – Theory and analysis volume 2* (pp. 974–1003). Berlin, München, Boston: De Gruyter Mouton.

Gass, S.M., Spinner, P., & Behney, J. (2018). *Salience in second language acquisition.* New York: Routledge.

Goldberg, A. (2006). Constructions at work: The nature of generalization in language. Oxford: Oxford University Press.

Goldberg, A., & Casenhiser, D. (2008). Construction learning and second language acquisition. In P. Robinson, & N. C. Ellis (Eds.), *Handbook of cognitive linguistics and second language acquisition (*pp. 197–215). New York: Routledge.

Goldschneider, J.M., & DeKeyser, R.M. (2005). Explaining the "natural order of L2 morpheme acquisition" in English: A meta-analysis of multiple determinants. *Language Learning*, 55, 27–77.

Gries, S., & Wulff, S. (2009). Psycholinguistic and corpus-linguistic evidence for L2 constructions. *Annual Review of Cognitive Linguistics*, 7, 163 – 186.

Han, Z., Park, E., & Combs, C. (2008). Textual enhancement of input: Issues and possibilities. *Applied Linguistics*, 29(4), 597–618.

Issa, B., & Morgan-Short, K. (2018). Effects of external and internal attentional manipulations on second language grammar development: An eye-tracking study. *Studies in Second Language Acquisition,* 41, 1–29.

Kusyk, M. & Sockett, G. (2012). From informal resource usage to incidental language acquisition: language uptake from online television viewing in English. *Asp, la revue dun GERAS*, 62, 45–65.

Lee, M. & Révész, A. (2018). Promoting grammatical development through textually enhanced captions: an eye-tracking study. *The Modern Language Journal*, 102(3), 557–77.

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.

Lee, S.K., & Huang, H.T. (2008). Visual input enhancement and grammar learning: A meta-analytic review. *Studies in Second Language Acquisition*, 30, 307–331.

Lenth, R., Buerkner, P., Herve, M., Love, J., Riebl, H., & Singmann, H. (2021). *Emmeans: Estimated marginal means, aka least-squares means // R package version 1.5.5-1*. Retrieved from https://cran.r-project.org/web/packages/emmeans/

Leow, R.P., & Martin, A. (2018). Enhancing the input to promote salience of the L2: A critical overview. In S. Gass, P. Spinner, & J. Behney (Eds.), *Salience in second language acquisition* (pp. 167–186). New York: Routledge.

Madlener, K. (2015). *Frequency effects in instructed second language acquisition*. Berlin, München, Boston: De Gruyter Mouton.

Majuddin, E., Siyanova-Chanturia, A., & Boers, F. (2021). Incidental acquisition of multiword expressions through audiovisual materials: The role of repetition and typographic enhancement. *Studies in Second Language Acquisition*, 1–24. https://doi.org/10.1017/S0272263121000036

Mayer, R. (Ed.). (2014). *The Cambridge handbook of multimedia learning*. Cambridge: Cambridge University Press.

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.

Muñoz, C. (2020). Boys like games and girls like movies. Age and gender differences in out-of-school contact with English. *RESLA*, 33, 172–202.

Muñoz, C., Pujadas, G., & Pattemore, A. (2021). Audio-visual input for learning L2 vocabulary and grammatical constructions. *Second Language Research*. https://doi.org/10.1177/02676583211015797

Olsson, N. L. (2019). Subtitle Edit [Computer Software]. nikse.dk. Retrieved from https://www.nikse.dk/SubtitleEdit

Paivio, A. (1986). *Mental Representations.* Oxford: Oxford University Press.

Pattemore, A., & Muñoz, C. (2020). Learning L2 constructions from captioned audio-visual exposure: The effect of learner-related factors. *System*, 93. https://doi.org/10.1016/j.system.2020.102303

Pellicer-Sánchez, A. (2017). Learning L2 collocations incidentally from reading. *Language Teaching Research*, 21(3), 381–402.

Pérez-Paredes, P., Mark, G., & O'Keeffe, A. (2020). The impact of usage-based approaches on second language learning and teaching. *Cambridge Education Research Reports*, 1–15.

Peters, E. (2019). The effect of imagery and on-screen text on foreign language vocabulary learning from audiovisual input. *TESOL Quarterly*, 53(4), 1008–1032.

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.

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.

R Core Team. (2020*). R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Retrieved from https://www.R-project.org

Robinson, P., Mackey, A., Gass, S.M., & Schmidt, R. (2012). Attention and awareness in second language acquisition. In S. Gass, & A. Mackey (Eds.), *The Routledge Handbook of Second Language Acquisition* (pp. 247–267). New York: Routledge.

Römer, U., & Garner, J.R. (2019). The development of verb constructions in spoken learner English: Tracing effects of usage and proficiency. *International Journal of Learner Corpus Research*, 5(2), 207–230.

Sharwood Smith, M. (1993). Input enhancement in instructed SLA. *Studies in Second Language Acquisition*, 15, 165–179.

Schur, M. (Creator). (2016). *The Good Place*. [TV series]. United States: Fremulon.

Taguchi, N. (2007). Chunk learning and the development of spoken discourse in a Japanese as a foreign language classroom. *Language Teaching Research*, 11(4), 433–457.

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.

Vanderplank, R. (2016). Captioned media in foreign language learning and teaching: Subtitles for the deaf and hard-of-hearing as tools for language learning. London: Palgrave Macmillan.

Webb, S. (2014). Extensive viewing: Language learning through watching television. In D. Nunan, & J.C. Richards (Eds.), *Language learning beyond the classroom (*pp. 159–168). New York: Routledge.

Webb, S., & Chang, A. (2015). How does prior word knowledge affect vocabulary learning progress in an extensive reading program? *Studies in Second Language Acquisition,* 37, 651–675.

Winke, P.M. (2013). The effects of input enhancement on grammar learning and comprehension. *Studies in Second Language Acquisition,* 35, 323–352.

Appendix A

A list of target constructions with frequency of occurrence

Construction type	Construction form and test type	Examples from *The Good Place*	Frequency
Fully-filled	do for a living[1]	What did you *do for a living*?	3
	let you down[1]	I won't *let you down*	3
	N[irregular plural] [3]	There are *shrimp* flying around	5
	big deal[1]	No *big deal*	6
	say no more[1]	Say no more	6
	figure out[2]	To *figure out* what's going wrong	11
Partially-filled	to be[tense] allowed to V[1]	I'*m not allowed to* tell you about	3
	would rather V[1]	I'*d rather* not *let* people see it	3
	break[tense] DET promise[2]	You *broke your promise*	3
	the Xer the Yer[3]	*The more* you practice, *the more* you improve	4
	used to V[1]	I used to just throw them in the sink	7
	PRON just want[tense] to[1]	*I just want to* be an academic	9
	let's V, shall we? [2]	So let's chat, *shall we*?	11
	why don't PRON[1]	*Why don't you* go ahead?	12
	to be[tense] supposed to V[1]	You *were supposed to* be there	18
	subj belong[tense] here[1]	You don't *belong here*	18
	let's V[2]	*Let's* move on	55
Fully-schematic	passive present continuous (subj aux VP) [1]	Her memory's still *being rebooted*	3
	future continuous (subj aux V-ing) [3]	Later this evening, we *will be enjoying*	3
	subjunctive (subj V that PRO V)[1]	You *wish* that you *were* related	4
	V[negative] either[2]	You're not supposed to be here *either*?	5
	passive present perfect (subj aux VP) [1]	It *has been proven*	14
	reported speech (reporting V (that) V) [1]	Tahani *said* that you *helped* Michael	15
	catenative V obj infinitive (sub V PRO to V) [1]	You *need me to* lie	19
	catenative V obj bare infinitive (let PRO V) [1]	Should I *let her stay*?	21
	future in the past (subj V[past] V) [1]	I thought transition *would* be easier	21
	emphasis (do[tense] V) [1]	That *does* sound like me	23

Notes on the test items. [1] Sentence transformation. [2]Fill-the-gap. [3] Complete the gap with a correct form of a word in brackets.

Appendix B

Examples of test items from the pretest and immediate/delayed posttests

I. Sentence transformation exercise:
*Complete each sentence with **two to five words**, including the word in **bold***

1. "We can help you with finding a flat," said my friends.
HELP My friends said _____ with finding a flat.

2. I hate it when people ask me what my job is because I am unemployed.
FOR I hate it when people ask me what _____because I am unemployed.

II. Complete the gap with a correct form of a given word:
Complete the sentences using a form of the words in brackets

3.The fisherman has sold about 500_____(shrimp) this morning.

4._____ (cold) it got, _____ (many) clothes they had to put on to keep warm.

III. Fill-the-gap exercise:
Complete the gaps with the appropriate word:

5. Let's go to the theater, _____ we?

6. You can't trust her, she always _____ her promises.

Appendix C

Immediate posttest. Constructions tested and test items.

	Constructions tested in the partial immediate posttests
Week 1	do for a living[1]
	let you down[1]
	let's V[2]
	passive present perfect[1]
	passive present continious[1]
Week 2	catenative V obj infinitive (sub V PRO to V) [1]
	emphasis (do[tense] V) [1]
	subj belong[tense] here[1]
	PRON just want[tense] to[1]
	let's V, shall we? [2]
Week 3	big deal[1]
	say no more[1]
	break[tense] DET promise[2]
	figure out[2]
	to be[tense] allowed to V[1]
	reported speech (reporting V (that) V) [1]
	future continuous (subj aux V-ing) [3]
Week 4	N[irregular plural] [3]
	to be[tense] supposed to V[1]
	used to V[1]
	future in the past (subj V[past] V) [1]
	the Xer the Yer[3]
Week 5	why don't PRO N[1]
	would rather V[1]
	subjunctive (subj V that PRO V)[1]
	V[negative] either[2]
	catenative V obj bare infinitive (let PRO V) [1]

Notes on the test items. [1] Sentence transformation. [2]Fill-the-gap. [3] Complete the gap with a correct form of a word in brackets

Appendix D

Example of a post-viewing activity

The purpose of this content comprehension multiple-choice activity was to keep students' attention on the episode and to integrate the viewing into the classroom. The comprehension activities neither included nor tested the target constructions.

Choose the correct answer to the questions about the episode that you have just watched.

1. How has Jianyu managed to stay undiscovered so far?
   a. He is very smart
   b. Tahani has been helping him
   c. He hasn't spoken a word

2. What is the name of the restaurant recently opened in The Good Place?
   a. The Good Plates
   b. Angel Cakes
   c. The Food Place

3. Why does Eleanor get no food in the restaurant?
   a. She is on a diet
   b. She is on a hunger strike
   c. She was on a hunger strike in the past

4. Why doesn't Eleanor want Jason Mendoza to be himself?
   a. She doesn't like his music
   b. She thinks she will be in trouble
   c. She thinks it will hurt Tahani's feelings

Appendix E

R packages and scripts

The *car* package (Fox *et al.*, 2020) with Anova() function was used to access the analysis of deviance, likelihood-ratio chisquare, and *p* values. The *emmeans* (Lenth *et al.*, 2021) package with emmeans() and pairs() functions was used to explore estimated marginal means and run the pairwise comparisons. Finally, the LMMs' effect sizes (marginal R^2 and conditional R^2) were calculated using *MuMIn* package (Barton, 2020).

Unconditional means model:

model1 = lmer (Test scores/maximum possible test score ~ (1 | Construction), data = dataset, REML = FALSE)

Research question 1a: Factors explaining construction learning:

model2 = lmer (Test scores/maximum possible test score ~ Group + Construction type + Time + Frequency + (1 | Construction), data = dataset, REML = FALSE)

Research question 1b: Interaction between construction learnability factors:

model3= lmer (Test scores/maximum possible test score ~ Group + Construction type + Frequency + Time + Group:Construction type + Group:Frequency + Group:Time + (1 | Construction), data = dataset, REML = FALSE)