UNIVERSITAT DE BARCELONA

Learning grammatical constructions from audio-visual input

Anastasiia Plotnikova

Doctoral Dissertation

Doctoral programme in Cognitive Science and Language

Learning grammatical constructions from audio-visual input

Anastasiia Plotnikova

Supervised by Dra. Carmen Muñoz Lahoz

UNIVERSITAT DE BARCELONA

2022

Barcelona

Abstract

This doctoral dissertation explores the effects of prolonged exposure to audio-visual input on learning a variety of English L2 grammatical constructions. The thesis is organised around three interconnected studies. The first study compares the effects of captioned and uncaptioned audio-visual input on learning grammar, and investigates whether this learning depends on such learner-related factors as L2 proficiency level, working memory capacity, and foreign language learning aptitude. The second study implements an additional captioning condition and compares grammar gains between the three captioning modes: Unenhanced captions, textually enhanced captions, and no captions. It also assesses the effects of three grammatical construction learnability factors: Construction type, frequency, and recency of occurrence. The third study turns to the viewers' perspective of learning from audio-visual input and addresses the participants' feeling of learning from the intervention, and whether their extramural L2 television viewing preferences and viewing strategies changed over the period of the intervention.

A total of 141 participants with various proficiency levels (from A1 to C2) watched ten full-length TV series episodes with captions, with textually enhanced captions (Study 2), or without captions over a period of five weeks. The study targeted 27 frequently occurring grammatical constructions categorized as fully-schematic, partially-filled, or fully-filled. The design included a pre-test, an immediate post-test, and a delayed post-test, along with pre-course and post-viewing questionnaires on participants' L2 viewing habits. Participants' individual differences such as English proficiency level, working memory capacity, and foreign language learning aptitude were also measured.

The results of the first study indicated that all participants significantly improved their knowledge of the target grammatical constructions, with the captions group having significantly higher gains than the no captions group. In terms of the individual differences, English proficiency had a mediating role with the intermediate level group outperforming the elementary group. The working memory capacity and language learning aptitude scores (LLAMA F) had a significant effect on the learning gains of the no captions group. Higher levels of these cognitive individual factors facilitated learning without captions, suggesting that captions can level the playing field while viewing L2 audio-visual materials.

The results of the second study 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. This finding suggests that in order to obtain effective learning conditions from textually enhanced captions, a limit to the amount of different textually enhanced constructions presented in the input should be set. In general, unenhanced captions - already available on the most streaming platforms – appear sufficient for successful grammatical constructions learning. As regards the type of constructions and frequency, fully-filled constructions were learnt the least by all the groups. It seems that constructions without any variation in the input were more difficult to uptake. Frequency of the target constructions occurrence did not have a significant effect on grammatical constructions learning of intermediate and advanced level participants. It is possible that lower language proficiency viewers are more sensitive to the frequencies in the audio-visual input.

The third study, exploring the learners' perspectives, found that the participants

perceived vocabulary and expressions as the most learnt language features. Concerning the preferred viewing mode outside of the classroom (with L1 subtitles, with L2 captions, without captions or subtitles), a significant shift was observed. All participants, regardless of their proficiency and intervention viewing group, opted for watching less with L2 captions. The elementary level participants found that viewing without any native language support was too challenging for leisure viewing, while the intermediate and advanced level students became familiar with the audio-visual input and turned into confident viewers. Finally, the participants also demonstrated a significant drop in applying viewing strategies, which could be attributed to them turning off L2 captions and becoming viewers rather than learners.

This doctoral dissertation extends the known benefits of sustained exposure to audio-visual input for L2 vocabulary learning and comprehension, and goes beyond by providing evidence for the potential of captioned L2 television for grammar learning.

Resum

Aquesta tesi doctoral explora els efectes de l'exposició prolongada a sèries de televisió en l'aprenentatge d'una varietat de construccions gramaticals de l'anglès com a llengua estrangera. La tesi s'organitza al voltant de tres estudis interconnectats. El primer estudi compara els efectes de les sèries amb subtítols en anglès (L2) i sense subtítols en l'aprenentatge de la gramàtica i investiga si aquest aprenentatge depèn de factors relacionats amb l'estudiant, com son el nivell de competència en la L2, la capacitat de memòria de treball i l'aptitud per a l'aprenentatge de llengües estrangeres. El segon estudi afegeix un tercer tipus de subtítols i compara els guanys gramaticals entre les tres condicions: subtítols no modificats, subtítols amb paraules ressaltades i sense subtítols. L'estudi també avalua els efectes de tres factors d'aprenentatge de la construcció gramatical: el tipus de construcció, la freqüència d'ocurrència i la proximitat d'ocurrència. El tercer estudi es centra la percepció dels espectadors sobre l'input audiovisuals i la sensació d'aprenentatge dels participants de la intervenció, i si les seves preferències i estratègies de visualització de televisió en la L2 fora de l'àmbit educatiu han canviat durant el període de la intervenció.

Un total de 141 participants amb diferents nivells de competència (de l'A1 al C2) van veure deu capítols complets de sèries de televisió amb subtítols, amb subtítols amb construccions ressaltades (segon estudi) o sense subtítols, durant un període de cinc setmanes. L'estudi es va centrar en 27 construccions gramaticals freqüents categoritzades com a totalment esquemàtiques, parcialment plenes o completament plenes. El disseny incloïa una prova inicial, i dues proves posteriors (una immediata i una posposada), juntament amb qüestionaris previs i posteriors a la intervenció, centrats en els hàbits de

consum de televisió en la L2 dels participants. També es van mesurar les diferències individuals dels participants, com el nivell de domini de l'anglès, la capacitat de memòria de treball i l'aptitud per a l'aprenentatge de llengües estrangeres.

Els resultats del primer estudi indiquen que tots els participants han incrementat significativament el seu coneixement de les construccions gramaticals, i que el grup de subtítols ha tingut guanys significativament més alts que el grup sense subtítols. Pel que fa a les diferències individuals, el nivell de competència en anglès té un paper mediador dins el grup de nivell intermedi, que supera el grup elemental. La capacitat de memòria de treball i l'aptitud per a l'aprenentatge de llengües (LLAMA F) tenen un efecte significatiu en els guanys d'aprenentatge del grup sense subtítols. Els nivells més alts d'aquests factors cognitius individuals faciliten l'aprenentatge sense subtítols, suggerint que els subtítols poden igualar les condicions per a tots els estudiants mentre es visualitzen materials audiovisuals en la L2.

Els resultats del segon estudi indiquen efectes mixtos dels diferents tipus de subtítols. Els subtítols amb construccions ressaltades (una condició més prominent) comporta resultats d'aprenentatge immediats, mentre que els subtítols no ressaltats tenen efectes més alts a llarg termini. Aquesta troballa suggereix que, per a obtenir condicions d'aprenentatge efectives a partir de subtítols ressaltats, s'hauria d'establir un límit a la quantitat de construccions ressaltades diferents que es presenten a l'input audiovisual. En general, els subtítols no ressaltats, ja disponibles a la majoria de plataformes de *streaming*, semblen suficients per a l'aprenentatge adequat de construccions gramaticals. Pel que fa al tipus de construccions i la freqüència, les construccions completament omplertes han estat les menys apreses en tots els grups. Sembla que les construccions sense cap variació en

l'input son més difícils d'aprendre. La freqüència d'ocurrència de les construccions no té un efecte significatiu en l'aprenentatge de les construccions gramaticals entre els participants de nivell intermedi i avançat. És possible que els espectadors amb una competència lingüística més baixa siguin més sensibles a les freqüències de l'input audiovisual.

El tercer estudi, que ha explorat les perspectives dels aprenents, ha trobat que el vocabulari i les expressions són percebuts com els elements més apresos pels participants de l'estudi. Pel que fa al mode de visualització preferit fora de l'aula (amb subtítols en català o castellà (L1), amb subtítols en L2 o sense subtítols), s'ha observat un canvi significatiu. Tots els participants, independentment de la seva competència i grup de visualització d'intervenció, han optat per veure menys televisió amb subtítols en la L2. Els participants del nivell elemental han trobat que veure sèries de televisió per plaer sense suport de la llengua materna era massa difícil, mentre que els estudiants de nivell intermedi i avançat s'han familiaritzar amb l'input audiovisual i s'han convertit en espectadors que se senten segurs. Finalment, els participants també mostren una davallada en l'aplicació d'estratègies de visualització, que es podria atribuir al fet que desactivaven els subtítols en la L2 i es convertiren en espectadors en lloc d'aprenents.

Aquesta tesi doctoral amplia els beneficis coneguts de l'exposició prolongada a l'input audiovisual per a l'aprenentatge del vocabulari i la comprensió global, i va més enllà proporcionant proves del potencial de la televisió en la L2 amb subtítols per a l'aprenentatge de la gramàtica.

Acknowledgments

This dissertation would not have seen its completion without a number of outstanding people on my way. First of all, the deepest thankfulness goes to my supervisor, Dr. Carme Muñoz, for always being there for me, constantly supporting me, and for being such a great role model and person. You have shown me how to enjoy academia and taught me so much throughout these seven years that I have known you. It seems that there is no way to thank you enough for everything you have done for me.

My enormous gratitude goes to Dr. Mar Suárez for generously allowing me to collect data in her classroom two years in a row. We have multiple projects to work on together and I am very glad to count on such an amazing co-author.

I would also like to thank my participants. Thank you for watching, enjoying, complaining, learning, and providing important insights. As a life-long language learner myself, I hope this experience has directed you into an additional source of language learning.

I am thankful to Oliver Valero for his statistical support. Thanks to you I can independently navigate through various statistical models.

I would also like to thank Dr. Teresa Cadierno, Dr. Geoffrey Sockett, and Dr. Roger Gilabert for providing constructive feedback on the initial PhD research proposal. Moreover, I would like to highlight the invaluable help of Dr. Cadierno with classifying the target constructions, thank you for dedicating several hours of your time in the middle of the busy lockdown.

The constant support from the GRAL members could not be overlooked, thank you for all the suggestions and feedback along the way, and thank you for treating young researchers as equals, this working environment is unique and precious.

I would also like to thank all the McDermott friends (Natasha, Valeria, Georgia, Ferran, Daniela, Natalia) for help with my data collection, coffee breaks, and weekend escapes from research. Also, thank you Geo for helping me with the Catalan version of the abstract.

Planning a research stay in the middle of the pandemic was not an easy task, and this leads me to another important person who I met on the way, Dr. Maribel Montero Perez. Thank you for accepting my last minute request to have a research stay with you, agreeing to collect data for a joint project, and being so supportive, kind, and generous to me.

Finally, we are nothing without our families. I am very thankful that my family taught me how to love learning and always encouraged me in my explorations of the world. It is a common belief that writing a PhD is a very lonely process, luckily I always have my husband, Matthew Pattemore, supporting me. I cannot thank you enough for being the best PhD buddy, reviewer, proof-reader, sense maker, IT consultant, and entertainer. I would not have survived this without you and I cannot wait to help you with your thesis submission later this year.

Thank you everyone for making my PhD experience so pleasurable, this has been an amazing *beginning* of my academic journey.

Table of contents

List of tables

List of figures

Chapter 1

Introduction

Exposure to language input plays a crucial role in language acquisition, and it has been shown that greater amounts of exposure to language input through activities such as watching TV, listening to music, and playing games results in higher second/foreign language (L2) proficiency - affecting L2 listening and reading comprehension (Lindgren & Muñoz, 2013). Audio-visual input, such as movies and TV series in the original version, has been proven a valuable resource for L2 development for both in-class and out-of-class exposure in different parts of the world (see Montero Perez, 2022). Moreover, it has been suggested that an extensive viewing approach, in which learners are exposed to larger amounts of input over time, could fulfil the need for exposure to ample amounts of L2 input (Webb, 2014). For several decades, researchers have also been exploring whether on-screen text could support this learning. For instance, it has been found that subtitles (translation of the L2 audiotrack into the first language (L1)) support comprehension (e.g. Pujadas & Muñoz, 2020), and that captions (the on-screen L2 text representation of the L2 audiotrack) facilitate learning of vocabulary and have a strong effect on content comprehension (e.g. Montero Perez, et al., 2013). While research on vocabulary learning has been proliferating, the number of studies on grammar learning through audio-visual input is scarce.

The grammatical focus of this study is based on the constructionist perspective of grammar, 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). Finally, 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). Out of those studies exploring L2 grammatical constructions, only two audio-visual studies have been conducted from a constructionist perspective of grammar, but neither looked specifically at the use of on-screen text. Kusyk and Sockett (2012) focused on incidental uptake of frequently occurring 4-gram constructions in five popular TV series. The results showed that the more time participants reported watching TV series, the better was their knowledge of the target constructions. Further research on regular watchers and L2 constructions use looked at fan fiction (Sockett & Kusyk, 2015). The results showed that compared to infrequent viewers, frequent viewers demonstrated significantly more use of target constructions in their written fan fiction production. Unfortunately, these studies did not explore effects of captioning on constructions learning from this type of input.

This doctoral dissertation focuses on the potential learning of grammatical constructions from prolonged viewing of captioned original version TV series, and comprises three interrelated studies. It starts with an overview of the current research on audio-visual input (Chapter 2), and distinguishes the promising areas of inquiry. The next chapter (Chapter 3) presents the aims and research questions of this dissertation. Chapter 4 provides a detailed study methodology to facilitate further (replication) studies and identifies participants, audio-visual materials used, target grammatical constructions, testing materials, and procedures. Chapter 5 reports on the first study of this dissertation that explores the effects of learner-related factors and of captions on learning grammatical constructions from audio-visual input. The mediating learner-related factors include viewer's proficiency, working memory capacity, and foreign language learning aptitude. The second study (Chapter 6) builds upon the findings of the first experiment and explores the linguistic-related factors along with different types of captioning mode: captions, textually enhanced captions, and no captions. The linguistic-related factors addressed in Chapter 6 consist of type of grammatical construction, individual frequency and recency of occurrence of the target constructions. The third study (Chapter 7) looks at the viewers' perspective of learning from audio-visual input and explores whether the participants perceived the intervention useful for their language development. It also investigates whether the participants' informal L2 television viewing preferences and viewing strategies had changed over the period of the intervention. Each individual study (Chapters 5, 6, 7) includes an introduction, study-specific methodology, data analysis, summary of the results, and discussion. The final chapter (Chapter 8) summarises the findings uncovered in the

three studies, addresses the studies' limitations, provides suggestions for further research, proposes how the findings could be applied to L2 teaching, and concludes this dissertation.

Chapter 2

Literature review

This chapter provides a comprehensive overview of previous research in the area of language learning through audio-visual input. It begins by describing the principles of audio-visual input and discusses how L2 learning from audio-visual input may be explained through the lens of information processing theories. Afterwards, it provides an outline of research done into various language dimensions including comprehension, vocabulary, pronunciation, and pragmatics, before arriving to the less frequently explored area of grammar learning. The sections on grammar learning through audio-visual input include studies incorporating various captioning modes. The chapter continues by addressing learner-related factors that might affect grammar learning from audio-visual input, including proficiency, working memory capacity, and language learning aptitude. It then follows by explaining the linguistic input-related factors affecting learning, and discusses such learnability factors as construction type, frequency, and recency. Finally, the last sections of this chapter describe research into learners' perspectives on the efficacy and use of audio-visual input.

Audio-visual input supported by captions has been a focus of attention of L2 research in recent years (see Montero Perez, 2022 for a comprehensive review of the audio-visual input research conducted within the last decade). Progressively, research is showing the

benefits of caption-supported audio-visual input for aspects of L2 learning, particularly vocabulary and comprehension (e.g. Montero Perez et al., 2013). Compared to vocabulary, scarce research has been conducted concerning grammar learning. Earlier studies focused on grammatical or morphological elements of language and have used subtitles in the L1 (d'Ydewalle & Van de Poel, 1999; Van Lommel et al., 2006); while more recent studies incorporated the comparison between unenhanced captions, enhanced captions, and no captions (Lee & Révész, 2018, 2020), and enhanced captions and no captions (e.g. Cintrón-Valentín et al., 2019).

2.1 Audio-visual input

Audio-visual input, often referred to as multimodal input, has been a focus of research for decades. The important feature of audio-visual input that differentiates it from other types of input (e.g. audio, textual), is its multimodality. Multimodality can be defined as a combination of "multiple different systems or signification to communicate a single, or at least unified, message or meaning" (Dressman, 2020: 39). An important characteristic of audio-visual input is that it simultaneously provides the input recipient with audio and with an image supporting it. Another crucial feature is the possibility of adding a third element into play – on-screen text that mirrors the audio. There are two main modalities of on-screen text: textual support in the same language as the audio, and textual support translated into a language other than the language of the audio. Although the terms differentiating these two types of on-screen text are not universal, this dissertation refers to them as captions and subtitles respectively. Subtitles, also known as native language (L1) subtitles, are a translation of the audio from the original language into another language – usually the

intended viewer's L1. This type of audio-visual input is common in a number of European countries that opt for the screening of original version television programmes, rather than dubbing (replacing the original audio with audio in a local language) (European Commission, 2011). Captions, also originally known as closed captions or captions for hard of hearing, are a simultaneous on-screen written text representation of the soundtrack in the same language as the audio. Captions may be used by L1 speakers of the input's original language to compensate poor hearing or audio quality, or by learners of that language to have both audio and text in the L2.

Various types of audio-visual input in L2 research have been explored: Static images with text and audio (e.g. Lee & Révész, 2018), educational videos specifically created for the experiments (e.g. Cintrón-Valentín et al., 2019), and, the focus of this dissertation, original version audio-visual input, such as documentaries (e.g. Teng, 2022), TED talks (e.g. Nguyen & Boers, 2018), YouTube videos (e.g. Lin, 2021), full-length movies (e.g. Vanderplank, 2019), and TV series (e.g. Rodgers, 2013).

Several researchers have argued that audio-visual input can stimulate L2 development. In his review of optimal input for language learning, Long (2020) suggested that audio-visual input with on-screen text can enhance incidental learning. In the same line, Vanderplank (2020) argued that exposure to audio-visual input, such as television, meets the requirement of Krashen's comprehensible input hypothesis (1985). Moreover, Webb (2014) has suggested that TV in the target language has the capacity to provide second language learners with large amounts of spoken input necessary for successful language learning. TV series – with a developing plot that encourages viewers to watch continuously – can be particularly helping in promoting an extensive viewing approach. Related to the

extensive reading approach (e.g. Nation, 2015), extensive viewing is defined as "regular silent uninterrupted viewing of L2 television inside and outside of the classroom" (Webb, 2014). Consequently, Rodgers (2016) connects the concept of extensive L2 television viewing to Nation's (2007) requirements for suitable L2 input. These conditions are that the input is processed in large quantities, and that the input is familiar to learners. Rodgers (2016) claims that TV viewing meets these conditions because L1 TV viewing is a common practice that could be transferred to L2 extensive viewing. He also suggests that the requirement of learners being able to learn from the input may be met due to the television programmes being rich in contextual cues – images, actions, and sound effects that may reinforce the meaning of the language being used. In addition, television also meets the criteria of input containing only a small amount of unknown vocabulary, as it was shown in Webb and Rodgers (2009) that knowledge of the 3000 most frequent words is sufficient to follow the content of a TV programme. Also, for input to be considered suitable, learners should be motivated and interested in engaging with it, and Rodgers (2016) advocates that the amount of different genres and lengths of the available television programmes allows audio-visual input to meet this criteria as well. Finally, he suggests that while L2 television already meets Nation's (2007) five criteria, the addition of captions may increase the suitability of extensive TV viewing as input even further.

In order to explain how learners process this type of input, research on L2 acquisition from audio-visual input has turned to processing theories which can provide appropriate theoretical foundations for language learning through videos. One of the relevant theories for audio-visual input is the Dual Coding Theory proposed by Paivio in 1986. This theory describes 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) asserts that people learn better when words are presented with pictures, as this allows learners to make connections between word and image.

Another central assumption of this theory is the limited capacity assumption which suggests that humans can only hold and process a limited amount of information in working memory (WM) at any given moment. Because the main information processing takes place in WM, it is here that the real complexities of multimedia come into play (Hede, 2002). Sweller's (2005) Cognitive Load Theory addresses the potential cognitive overload resulting from the fact that the amount of information able to be processed at one time in each channel (verbal or visual) is limited. It suggests that the presentation of the same information in different forms (i.e., oral and written) may result in redundancy, leading to a split-attention effect and lower learning gains (Ayres & Sweller, 2014). This is in line with the redundancy principle (Mayer, 2009) that proposes that viewers benefit from multimodal input more when they are presented with image and audio, rather than with image, audio, and on-screen transcription of the audio.

However, the above-mentioned negative effects of tri-modal input were found to be related to multimodal input in the viewer's L1, and appear to be overruled when these principles are applied to L2 viewers (Mayer et al., 2020). In order to address these contradictory effects of multimodal input in L1 and L2, Mayer and colleagues (2020) proposed the subtitle principle: That on-screen text provides L2 viewers an opportunity to revisit the language content if they were unable to fully process the audio. This way, viewers would have a 'backup option' and their working memory would not become overloaded.

Similarly, Vanderplank's cognitive-affective model of language learning through captions viewing (2016) claims that the addition of captions benefits language learners. This is because the information is distributed between the three input sources (audio, visual, and caption text), allowing them to complement each other and support learning. Rather than overloading the viewer's cognitive capacity, captions would maximise L2 learning from audio-visual input, as the three input sources balance each other's processing loads (Frumuselu et al., 2015).

2.2 L2 learning from audio-visual input

A number of studies have explored the effects of audio-visual input on language learning and found that videos, with or without on-screen text (i.e. subtitles or captions), may have effects on various language features. This section provides an overview of research exploring the effects of original-version audio-visual input on L2 comprehension, vocabulary, speech segmentation, and pragmatics.

Interest in promoting language learning through audio-visual materials has been increasing through the last decades with several international symposia and workshops dedicated to the topic (e.g. EuroSLA/Language Learning symposium on multimodal input 2021), and special issues in top-tier journals (e.g. *The Language Learning Journal* (Montero Perez & Rodgers, 2019), *Studies in Second Language Acquisition* (Muñoz & Peters, 2020)). There have also been several comprehensive review publications covering decades of research on audio-visual input (Caruana, 2021; Montero Perez, 2022; Vanderplank, 2010; Vanderplank, 2016). Due to the great amount of studies on audio-visual input, and the scope

of this doctoral dissertation, only studies analysing effects of original-version films and television series will be discussed in this section.

Listening comprehension and vocabulary learning are the most explored features in audio-visual input research. A meta-analysis on the effectiveness of captioned audio-visual input found a large effect of captions on comprehension and vocabulary learning (Montero Perez et al., 2013). In L2 comprehension studies, subtitles have been found to be more effective for L2 comprehension than captions (Fievez et al., 2020; Pujadas & Muñoz, 2020). In a study comparing captioned and uncaptioned viewing of ten TV series episodes (Rodgers & Webb, 2017) both intervention groups significantly improved their comprehension scores over the intervention time period. However, the results suggested that the captions group performed better than the no captions at the beginning of the experiment, suggesting that the no captions group needed some time to tune in. It was also found that captions facilitated comprehension when the episodes were particularly difficult.

As for vocabulary studies, Rodgers and Webb (2020) found that significant incidental vocabulary learning with a large effect size can appear after viewing ten episodes of a TV series (approximately 420 minutes) without captions. Similarly, Peters and Webb (2018) encountered substantial incidental vocabulary learning after viewing a one-hour documentary without captions. As for the difference between types of on-screen texts, studies have found a positive effect of viewing under different conditions with some advantage for captions over subtitles (Fievez et al., 2020; Peters, 2019; Pujadas & Muñoz, 2019), and for captions over no captions (Montero Perez et al., 2014; Peters, 2019; Sydorenko, 2010).

Apart from single words, research has also been exploring the learning of multi-word units (e.g. Majuddin et al., 2021) and formulaic sequences (e.g. Puimège & Peters, 2020) from audio-visual input. To give an example, a study by Puimège and Peters (2020) explored the learning of formulaic sequences from a one-hour documentary without captions or subtitles. The upper-intermediate English proficiency level participants completed form recall and meaning recall pre-/ and post-tests. The results indicated that incidental uptake of formulaic sequences can happen from viewing an hour of uncaptioned audio-visual material, and that on average ten formulaic sequences were learnt in the form recall test, and seven in the meaning recall test. Another study focusing on learning colloquial expressions from audio-visual materials compared the effects of two types of on-screen text: L1 subtitles (Spanish) and L2 captions (English) (Frumuselu et al., 2015). The elementary to advanced level proficiency participants watched 13 episodes of a TV series (approximately 325 minutes), and completed meaning recognition and recall pre-/ and post-tests. The results yielded a significant advantage of L2 captions over L1 subtitles, suggesting a benefit of watching captioned audio-visual materials for informal expressions learning.

Vocabulary studies have also started going beyond standard captioning and subtitling, and explored such on-screen text modes as bilingual captions (e.g. Liao et al., 2020; Wang & Pellicer-Sanchez, 2022) and glossed captions (e.g. Fievez et al., 2021). In the study by Wang and Pellicer-Sanchez (2022), the processing of dual subtitles, a condition when the on-screen text is presented in the viewer's L1 and L2 simultaneously, was explored along with the vocabulary gains. The participants watched a 23-minute documentary under four different condition: Dual subtitles in Chinese and English, captions in English, subtitles in Chinese, and no on-screen text. It was found that captions were better

for learning vocabulary forms, while bilingual subtitles were beneficial for vocabulary meaning recall and recognition. However, the authors suggested that Chinese viewers are already familiar with this type of captioning; therefore it would be relevant to compare the promising results with other language pairs where this captioning style is less familiar.

Advances in technology are providing further possibilities for language learning through textually supported audio-visual input. An example of such advancement is the free Chrome extension "*Language Reactor"* (Wilkinson & Apic, 2018) that allows viewers to highlight the target words throughout the video, activate dual subtitles, and get access to glossed captions. Fievez et al. (2021) explored the potential of this tool for vocabulary learning by analysing glossed captions. Glossed captions allow viewers to access a word's L1 translation and examples of its use if they click on a word in the caption line. The low to high-intermediate proficiency participants watched six episodes of an L2 French TV series (a full season, 307 minutes) on *Netflix* outside of the classroom. Significant form and meaning recall gains appeared after the extensive viewing. The number of times the participants clicked on the target words to see the L1 translation was positively correlated with meaning recall scores, and a slightly lower effect on the form recall gains, suggesting that glossed captions are particularly beneficial for learning the meaning of the words. Although the intervention led to significant vocabulary learning, not all the students fully engaged with the glossed captions. Some participants shared that they did not have enough time to access the translations due to the limited time the captions appear on the screen. Other participants suggested that they realised they did not need to understand each and every word, and therefore their use of glosses was limited. Further research should explore

this promising area of captioning and compare the effects of traditional and glossed captions with this tool.

Due to audio-visual input containing spoken language in context it has also raised interest in research focused on L2 speech in order to see whether captions help with decoding aural messages. The studies discussed below used a shadowing task to measure speech decoding skills; in this task participants are asked to listen to a phrase and repeat it as accurately as possible. Mitterer and McQueen (2009) explored whether viewing a 25-minute video with captions, subtitles, or no on-screen text could improve highly advanced English learners' perception of strongly accented speech. The researchers found that captions helped with L2 speech decoding and outperformed both subtitles and no on-screen text groups. Charles and Trenkic (2015) analysed participants' speech segmentation after viewing two 30-minute documentaries under three conditions: with captions, without captions, and with captions but without audio. The results indicated that captions (with audio) not only assisted with speech segmentation of the phrases that were present in the input better than the rest of the groups, but also that experience with viewing audio-visual input with captions further led to better decoding of unfamiliar phrases that were presented only auditorily. Another study explored the effects of audio-visual input on L2 speech segmentation after a prolonged exposure (almost 6 hours) to one season of a TV series (Wisniewska & Mora, 2020). The participants watched the episodes with or without captions, and the results indicated that extensive exposure to TV series improved participants speech processing skills regardless of the viewing condition. The results seem not to show the previously found advantage of the captioned over the uncaptioned condition,

possibly due to the longer period of exposure to the audio-visual input when compared to prior studies.

The last language feature to be discussed in this section is pragmatics, one of the areas that is only starting to be explored through the audio-visual input perspective. According to Martínez-Flor (2008), audio-visual materials have the potential of being useful to introduce various pragmatic aspects in a classroom. The researcher analysed ten English films and found that movies were a valuable source of different request modification devices, compared to L2 English coursebooks that did not have adequate instances of requests (Fernández Guerra & Martínez-Flor, 2003). Martínez-Flor (2008) advocated that authentic audio-visual materials can bring real language into the classroom and provide meaningful pragmatic input. Additionally, Usó-Juan and Martínez-Flor (2021) suggested that the use of films in a classroom may raise students' (meta)pragmatic awareness.

Following these ideas of the potential of audio-visual input for L2 pragmatics learning, Khazdouzian et al. (2021) explored learning of request and suggestion strategies from a whole season of a TV series with or without captions (24 episodes, 516 minutes). The researchers observed some improvement in pragmatic strategies, but this could not be attributed to on-screen text, as the captions group did not produce request and suggestion types that had been present in the input. The authors suggested that the weak captioning effect could be due to the participants viewing the episodes at home as a leisure activity, and there was no instruction or focus on pragmatic aspects. In a recent study, Barón and Celaya (2022), examined production of pragmatic features after viewing seven five-minute excerpts from different TV series twice with captions or without captions. The results confirmed earlier findings in the field that audio-visual input was beneficial for L2

pragmatic development regardless of viewing condition, but it was also found that the captions group used more pragmatic features that appeared in the excerpts. In another study, pragmatic-focused instruction was considered as a way to support learning from captioned audio-visual input by intermediate level learners (Barón & Muñoz, in press). The study focused on greetings and requests, and compared the effects of captioned audio-visual input for two age groups: Children from 11 to 13 years old, and adults from 30 to 50 years old. Both age groups were divided into two sub-groups: instruction, viewing, and practice (intervention group); and instruction and practice (control group). The treatment included six 20-minute episodes with captions that were viewed on separate days. The results of the post-test indicated that the adult intervention group outperformed the rest of the groups, and there was a tendency for the child intervention group to perform better than the child control group. The authors proposed that adults benefitted more from the audio-visual materials due to already having a developed L1 pragmatic competence that affected their L2. Overall, the results showed the potential of captioned TV series to support explicit pragmatic instruction.

The L2 pragmatics studies in the area of audio-visual input indicate that TV series may bolster pragmatic expressions, but more research is needed to establish what type of on-screen text is beneficial for learning and producing pragmatically appropriate utterances.

To summarise this section, the extensive research on the effects of audio-visual input and language learning has been focused on various language features – comprehension, vocabulary, speech decoding, and pragmatics - and has investigated factors such as length of exposure and captioning mode. The research is not limited to these topics, as there are also studies exploring learner-related factors: proficiency (e.g. Suárez & Gesa, 2019), working memory capacity (e.g. Montero Perez, 2020), and aptitude (e.g. Suárez & Gesa,

2019); and input-related factors: frequency (e.g. Muñoz et al., 2021), imagery (e.g. Peters, 2019), and repeated viewing (e.g. Majuddin et al., 2021), among many others. The next section discusses studies on a less explored language feature, and the focus of this doctoral dissertation: grammar.

2.3 Grammar learning through audio-visual input

Very few pieces of research have specifically addressed grammar learning through audio-visual input, and even fewer through original version materials. Earlier studies focusing on grammatical or morphological elements of language used subtitles in the L1. For example, d'Ydewalle and Van de Poel (1999) explored young learners' French and Danish syntax and morphology learning under two subtitling conditions: reversed (L1 audio, L2 subtitles) and standard (L2 audio, L1 subtitles). The results showed no significant grammar gains for either of the languages or conditions. In the same line, Van Lommel et al. (2006) studied a potential effect of pre-teaching on grammar learning from subtitled materials. The study comprised two experiments each with a different subtitling condition: a 25-minute video with reversed subtitles (L1 Dutch audio, L2 Esperanto subtitles), and a 16-minute video with standard subtitles (L2 Esperanto audio, L1 Dutch subtitles). The results indicated that better performance in the post-test was associated with grammar pre-teaching before the viewing, suggesting that, without any rule presentation, a single exposure to audio-visual materials may not lead to significant grammar uptake. The authors concluded that, compared to vocabulary, grammar may be too complex to pick up incidentally from a single video, and that grammar learning from audio-visual input may require additional instruction and practice. The authors also hypothesised that language

learners probably start from acquiring chunks of language, and only after they receive substantial amount of exposure to audio-visual materials will they extract grammar rules from the input. Similarly, other scholars have speculated that captions may have the potential to lead to grammar uptake only after a prolonged exposure to audio-visual input (e.g. Kuppens 2010; Matielo et al., 2015).

More recent studies on grammar learning have explored the potential effects of captions, rather than native language subtitles. For instance, my Master's thesis (Plotnikova, 2017) analysed the learning of English conditionals from one TV series episode. A short intervention of repeated 14-minute viewings of the same captioned video led to a significantly higher performance in the post-test (as compared to the pre-test), especially for learners with higher proficiency level. However, the absence of a no captions group means it was not possible to claim that the learning could be attributed to captions.

Further audio-visual input studies focusing on grammar have mostly incorporated a comparison between different types of captioning. The next section presents another type of captioning – textual enhancement - and provides an overview of the grammar studies exploring the effects of various captioning modes.

2.4 Textually enhanced captions

Different captioning modes and their effects on learning gains have been the subject of recent research. These captioning modes include full captions, key-word captions, textually enhanced (TE) captions, and no captions. TE captions, in particular, have attracted 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 captions, captions with highlighted key-words, and key-word captions) 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 (underlined) (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 post-test 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 post-test 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).

Another study on learning multiword units from textually enhanced captions implemented an eye-tracking methodology to uncover the processing of textually enhanced captions and their effects on learning (Puimège et al., 2021). The participants viewed a 30 minute documentary while their eye-gaze movements were recorded. The within-participants study compared the effects of textually enhanced (underlined) and unenhanced captions on learning 28 target multiword units that appeared once each throughout the video. The counterbalanced design meant that some participants watched the first half of the video with TE captions, and others watched the second half with the TE captions. This allowed the researchers to compare the effects of enhanced and unenhanced captions on learning target items within the same participant pool. The pre-/post-tests included a form recall test. The results indicated that the enhanced captions attracted viewers' attention and led to more rereading. Although there was a positive relationship between the time spent on the TE captions and the immediate post-test results, the effect of textual enhancement disappeared once total reading time was included as a factor. This suggested that TE captions can successfully attract viewers' attention, but do not necessarily lead to higher learning gains. The authors suggested that general engagement with captions seemed to be a stronger predictor of learning, at least for multiword units that are presented in audio-visual materials. They proposed that the underwhelming findings regarding enhanced captions

could be due to the nature of audio-visual input where the viewers do not have a lot of time to reread the textually enhanced captions and, consequently, process them at a deeper level.

2.4.1 Grammar and textual enhancement

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 (Indrarathne et al., 2018; 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). Regarding the cognitive processing of textually enhanced grammatical constructions, Indrarathne and colleagues (2018) implemented an eye-tracking methodology to measure the changes in the processing of a causative *had* construction (*had something done*) in written texts. The intermediate level learners of English were exposed to three different texts that included seven instances of the target construction (21 in total). The results indicated an attenuation in the attention paid to the enhanced target constructions (in bold) as the participants encountered it in the input more times. The authors suggested that attention to the target construction decreased as it became less novel.

In the area of audio-visual input five studies have focused on the effectiveness of textual enhancement on grammar learning from static images and short clips. Lee and

Révész (2018) compared the effects of TE captions (in bold) with non-enhanced captions for learning of L2 grammatical constructions (third-person pronominal anaphora reference) 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 post-test 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 brief animated educational 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 pre-test 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. In addition, as the pre-teaching resulted to be a significant factor in the post-test grammar scores, it is complicated to determine whether the gains in the two studies appeared due to the exposure to audio-visual input or effective pre-teaching. Concluding this series of studies on learning Spanish grammar from a multimodal intervention, their latest study with the same brief animated videos explored the effects of two different enhancement techniques when compared to uncaptioned clips (García-Amaya & Cintrón-Valentín, 2021). The two TE techniques included enhancement of the whole grammar structure in bold and yellow, and enhancement of the inflectional morphemes only. The target structures included gustar-type verbs,

preterite-imperfect contrast, and subjunctive. Interestingly, in the inflectional morpheme enhancement condition of subjunctive grammar, there was also an arrow to demonstrate which verb caused the subjunctive mood. The participants were familiar with the target grammar as they received explicit instruction on these structures in their usual L2 Spanish classroom (pre-teaching was not a part of the experiment). The results of the written elicited imitation test that was taken during the viewing of the videos indicated that both TE groups outperformed the no captions group in all target features. As for the difference between the two TE techniques, the subjunctive scores (with the arrows) were significantly higher in the inflectional morphemes TE group. Although the results indicated an advantage of TE captions over no captions, the study also did not include an unenhanced captions condition, making it impossible to tease apart the effects of captioning and textual enhancement.

To summarize, research on grammar learning from various captioning modes is limited to studies that provided only short[1] 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

[1] We refer to these exposures as "short" based on the length of videos reported in the studies, except for Cintrón-Valentín et al. (2019), Cintrón-Valentín and García-Amaya (2021), and García-Amaya and Cintrón-Valentín (2021) who did not report the length of their clips. However, on the basis of the two full scripts of the videos provided by the authors, they must have been shorter than 5 minutes.

et al., 2019). Therefore, the learnability factors of grammatical constructions may help clarify the inconsistent findings in this area (Section 2.6.).

2.5 Learner-related factors

This section addresses the individual factors that might affect learning from audio-visual input and includes language proficiency, working memory capacity, and language learning aptitude.

2.5.1 Language proficiency

The potential moderating effects of an individual's L2 proficiency on learning from audio-visual material has been frequently discussed (e.g., Vanderplank, 2016:80). Webb and Rodgers (2009) analysed the vocabulary demands of English television programmes and found that in order to follow the content of a film or a TV series, the viewer needs a knowledge of around 3000 word families, approximately equivalent to the intermediate proficiency level (Milton & Alexiou, 2009).

As for the interaction between proficiency and captioning, it has been argued that captions might not be as valuable for beginners as they are for more advanced learners due to beginner level students' limited processing capacity in the target language (e.g., Danan, 2004; Taylor, 2005). For instance, Danan (2004) proposed a minimum language competency threshold for language learners to benefit from captioning. Consequently, she suggested that the audio-visual materials should be carefully prepared for each level of language proficiency to provide comprehensible input which would result in beneficial learning outcomes. In another study, Taylor (2005) suggested that beginner level learners

may have limited processing capacity in L2 and difficulties in paying attention to three channels of input (audio, picture, and text) simultaneously. The author suggested that captions may assist beginner level learners only after they become familiar with this type of input. In contrast, some authors have found evidence supporting the view that even beginner learners might benefit from captions, either key-word or full (e.g., Guillory, 1998; Markham, 1989, respectively). More recently, Mirzaei et al. (2017) showed the effectiveness of captions for low-proficiency learners' listening skills in a study in which the researchers adjusted the amount of textual information in the on-screen text to match the learner's level. Another study with low-proficiency learners (Teng, 2019) compared the effects of different types of captioning (full captions, key-word captions, and no captions) on listening comprehension of a ten-minute video. The participants were elementary level primary school learners and were divided into low and high proficiency groups based on their Flyers Cambridge Exam scores (an A2 level exam). The results yielded that higher proficiency students outperformed the lower proficiency peers in all three conditions.

Several studies have shown significant learning by all participants, regardless of their proficiency, yet, with higher proficiency learners obtaining higher gains. For example, for vocabulary (Pujadas & Muñoz, 2019) adolescent learners with higher proficiency (pre-A level or A2/B1 level) outperformed their classmates after extensive exposure to captioned TV series. In the context of young adult learners with varying levels of proficiency (from A2 to C1), Suárez and Gesa (2019) found that higher proficiency predicted better vocabulary scores after a sustained exposure to captioned TV series. In a recent vocabulary study, Gesa and Miralpeix (2022) explored learning by various proficiency groups from 8 episodes of the TV series. The three proficiency groups included elementary (primary

school), intermediate (secondary school), and high-intermediate (university students). The elementary group watched the TV series with L1 subtitles (Spanish), and the intermediate and high-intermediate groups watched the TV series with L2 captions (English). In the treatment groups the participants received explicit vocabulary instruction before viewing the episodes, and in the control groups the participants only received traditional teaching. The results indicated that the elementary group did not differ from the control group, suggesting that L1 subtitles may have hindered vocabulary learning from audio-visual input. The intermediate group who viewed the episodes significantly outperformed the control group, and there was no difference between the viewing and the instruction only groups in the high-intermediate level. The authors suggested that for the high-intermediate level traditional teaching was enough to grasp the forms and meanings of the target words, while intermediate proficiency seemed to gain the most benefits for learning from audio-visual materials support.

Studies also found that vocabulary size plays a major role in learning from captioned audio-visual input, with the learners with bigger vocabulary sizes reaching higher vocabulary (e.g. Suárez et al., 2021) or formulaic sequences (e.g. Puimège & Peters, 2020) gains. However, the findings that vocabulary size affects vocabulary uptake are not consistent throughout different studies, especially those with extensive exposure to audio-visual materials where little or no effect of viewers' vocabulary size was observed (Fievez et al., 2020; Rodgers, 2013). Finally, for grammar, Plotnikova (2017) explored learning from repeated exposure to one captioned extract of a TV series (28 minutes in total). The participants were B1-B2 level adolescents. Three of the target grammar structures were pre-taught, and one grammar structure was not, comprising the incidental learning condition.

The results showed a tendency for the higher proficiency students to score higher for the pre-taught structures, and a significant effect of higher proficiency level for the not pre-taught structure, suggesting an advantage of higher proficiency for learning from captioned audio-visual input, especially for incidental learning.

Contrarily, Winke et al. (2010) compared the results of second and fourth year university learners of Spanish and Russian and found that their participants benefited from listening activities provided by a captioned video intervention equally, regardless of proficiency level. They suggested that captions do not have different effects on various proficiency levels. Similarly, in their meta-analysis of the studies on listening comprehension and vocabulary learning from captioned audio-visual input, Montero Perez et al. (2013) suggested that captioning may be an effective tool regardless of proficiency level as long as the audio-visual materials are adequate for the proficiency level of a viewer.

While proficiency appears to play an important role in learning from audio-visual input, the mixed results in the field suggest that more research is needed to shed light on this variable and particularly on its effects on grammar learning, which has received less attention in the audio-visual input research. It is also important to determine what captioning mode (e.g. captions or no captions) is more beneficial for various proficiency levels, as the majority of the studies did not take the proficiency and captioning mode interaction into account.

2.5.2 Working memory capacity

Another individual difference that has started attracting audio-visual input research is working memory (WM) capacity. Baddeley (2012) defined the multi-componential model

of WM as a system for keeping information in mind while performing a variety of complex cognitive activities, such as learning, problem-solving, language comprehension, and language production. WM consists of an executive processor and two short-term stores: a visuospatial sketchpad and a phonological loop. Multimedia researchers have been accepting the importance of WM and investigating the factors that may affect the processing of multimedia information in WM storage: cognitive overload and dual-coding. To give an example, transitory auditory information (e.g. speech) may entail a greater memory burden than information coming from other channels, such as textual for example (Xu et al., 2008).

Research on the relationship between L2 learning outcomes and individual differences in WM capacity has uncovered a strong positive correlation (see the meta-analyses by Linck et al., 2013; Tagarelli et al., 2015). Several studies have explored the relationship between grammar learning and WM capacity, for instance, Martin and Ellis (2012) claimed that compared to vocabulary learning, the acquisition and use of grammar requires greater processing because it not only requires memorizing items, but also abstracting them as a whole or as patterns. These complex grammatical items demand more cognitive processing and, consequently, their acquisition may be more predisposed to cognitive individual differences in WM capacity. The authors' results confirmed that claim and showed that WM had a strong relationship with participants' ability to generalize and apply novel grammar rules of an artificial language in both production and comprehension tasks. McDonough and Trofimovich (2016) looked at both artificial language and L2 English and found no contribution of WM to grammar learning. The authors claimed that the task may not have been challenging enough to activate the WM storage needed to perform it, and that WM might be more relevant in more explicit, intentional, and instructed

learning conditions. Indrarathne and Kormos (2017) explored the effects of WM capacity on explicit and implicit learning of unknown grammatical constructions. The results showed a clear advantage for higher WM regardless of the experimental condition. All language learners with higher WM performed better than the lower WM students, suggesting that the latter group needed extra support while acquiring previously unknown grammatical constructions.

WM capacity is likely to be an important factor in multimedia learning because students have to attend to multiple pieces of information coming from various sources and channels. According to Wiley et al. (2014), a major benefit of using multimedia instruction for learning is that it can neutralise WM limitations. More importantly, learners with a lower WM capacity may need supplementary support, such as explicit instructions to focus on a particular form, in order to construct knowledge while being exposed to multimedia materials.

The scarce research on audio-visual input that has explored the effects of WM capacity has yielded mixed results. For instance, Matielo et al. (2018) explored the interaction between WM capacity, L2 vocabulary learning, and content comprehension. The participants watched one 20-minute episode of a TV series under one of the viewing conditions: With captions, with L1 Portuguese subtitles, and without on-screen text. Different conditions were not significantly affected by differences in WM capacity – neither for vocabulary, nor for comprehension. It is possible that the small sample size of only 12 participants per group may have prevented the results from reaching significance. Later, Montero Perez (2020) looked at the role of WM on acquisition of L2 pseudowords through audio-visual input. The participants watched a 45-minute L2 French documentary without

captions and completed several vocabulary post-tests along with complex WM tasks. Higher WM scores contributed to higher scores in vocabulary recognition tests, indicating that individual differences in WM capacity can predict learning of novel vocabulary from uncaptioned video viewing. However, this study did not include a comparison group with captions to see whether differences in the WM capacity are affected by various viewing modes. In case of captioning and WM capacity, Gass et al. (2019) investigated the effects of WM capacity on L2 comprehension and reading behaviour of captions. The participants watched a 4-minute 37-second video twice in L2 English (English as a Second Language (ESL) learners) or Spanish (Foreign Language (FL) learners). The results showed no significant variation between the experimental groups, but there was a tendency for ESL participants with higher WM capacity to perform better in a comprehension task. The researchers explained this tendency between the groups by suggesting a proficiency threshold needed to activate WM processing. Although participants' proficiency was not measured, the ESL group was believed to have a higher language proficiency than the FL group. The researchers proposed that participants with higher proficiency obtained the key content information from the first viewing and, of those, the ones with higher WM could successfully keep it in the episodic buffer of their WM. Consequently, during the second viewing they could focus on the details of the documentary, rather than concentrating on the captions. The researchers proposed a possible neutralising effect of captions where captions help to counter-balance differences in WM capacity. In a more recent study on the effects of vocabulary size, attention, inhibition, and working memory (Suárez et al., 2021) the researchers explored the effects of captioned videos on early vocabulary learning by the means of form and meaning recognition tests. While the results suggested that the learners'

vocabulary size (proficiency) was the main predictor of learning from the exposure to four short clips (about 4 minutes each), there was no mediating effect of the participants' WM. The authors suggested that viewing short captioned clips was not complex enough to activate the effects of WM. Another study implementing short clips used treatment video tasks and explored the effects of the WM on learning the present perfect structure from clips (Lee & Révész, 2021). The participants were advanced learners of English allocated to one of three viewing conditions: with captions, with TE captions, no captions. They watched 24 news clips (20-50 seconds) and completed oral production, written production, and fill-in-the-blanks tasks. Working memory tests included phonological short-term memory and visual short-term memory measurements. The results showed that phonological short-term memory had a positive correlation with oral production gains in the captioned group, and with oral and written production gains in the TE captions group. Interestingly, participants in the no captions group relied on a different construct of the WM: Visual short-term memory. There was a positive relationship between the visual short-term memory and oral production gains in the no captions group. Overall, the results suggest a mediating role of working memory on learning from video tasks.

Working memory is sometimes included in a construct of general aptitude (Wen et al., 2016) and even considered the central component of the foreign language aptitude construct by some researchers (Miyake & Friedman, 1998). Nevertheless, there are reasons for distinguishing these two constructs because of the lack of consistent findings concerning a strong correlation between measures of WM and measures of aptitude (e.g., Hummel, 2009; Kormos, 2013). Additionally, principle component analysis studies have showed that WM and aptitude components load on different factors (Granena, 2013; Yalçın et al., 2016).

The next section discusses the role of foreign language learning aptitude in language learning from audio-visual input.

2.5.3 Foreign language learning aptitude

Language learning aptitude is a specific talent or a set of abilities which predicts capacity, readiness, rate, and speed of language learning (Wen et al., 2016). It has also been regarded as one of the strongest predictors of rate of language learning and ultimate L2 attainment (Granena & Long, 2013). The framework of FL aptitude initially introduced by Carroll in the 1980s and adopted by more recent proposals such as Meara's LLAMA battery of tests (2005) remains the most influential in SLA research (Wen et al., 2016). Although language learning aptitude can be seen as a "monolithic construct", it is also warranted to analyse language learning aptitude componentially (Skehan, 2002). For instance, Carroll (1990) proposed four aptitude constructs, consisting of phonemic coding ability, grammatical sensitivity, inductive language learning ability, and associative memory. Similarly, the LLAMA tests also measure four constructs or abilities: Ability to learn new words, ability to recognise words in the auditory input, ability to make a connection between a sound and a symbol, and ability to learn new grammar rules (Meara & Rodgers, 2019).

The most frequently reported feature to be correlated with FL aptitude is L2 grammar. A meta-analysis of the role of FL aptitude on L2 grammar learning (Li, 2015) reported results from 33 studies focused on the acquisition of morphosyntax. Although it was found that aptitude scores were moderately associated with grammar learning, the results revealed an overestimation of its role. First, FL aptitude was more likely to predict the rate of acquisition at initial stages of learning and not at later stages. In line with these

findings, Winke (2013) found that FL aptitude was not particularly important for her advanced learners of Chinese, implying that language aptitude plays a more important role with lower-proficiency learners. In fact, some studies have confirmed the strong role of FL aptitude in young school learners with a beginner proficiency level (Muñoz, 2014; Suárez, 2010). Therefore, it is reasonable to predict that the learning process of the lower proficiency learners would be more affected by FL aptitude.

Another finding in Li's (2015) meta-analysis suggested that FL aptitude was more likely to affect learning outcomes in explicit rather than implicit conditions. Some researchers have seemed to verify this claim (Granena, 2013), though it has not necessarily been confirmed by later research (Yalçın et al., 2016).

There has been a growing focus on the task-dependent nature of language learning aptitude in SLA (Granena, 2013; Kormos, 2013; Robinson, 2012; Skehan, 2002). The results of these studies have provided evidence that learners rely on different aptitude components in response to different L2 processes (see Yalçın et al., 2016). Consequently, it may be suggested that language aptitude, globally or componentially, may play a distinguishing role in processing and attending to the multiple pieces of information coming from three audio-visual input channels: audio, visual, and textual. However, research on audio-visual input that takes FL aptitude into account is restricted to three studies. Those studies explored the effect of language aptitude measured by the LLAMA battery of tests (Meara, 2005) on learning vocabulary from audio-visual materials.

Suárez and Gesa (2019) explored vocabulary learning of upper-intermediate level undergraduate students watching eight captioned episodes of a TV series (approximately 24 minutes per episode). The authors found that language aptitude did not explain the students'

gains in word forms, but it had a significant impact on their meaning recognition of the target words. They suggested that aptitude did not influence word form learning because the learners' focus of attention was more explicitly drawn to target words' meaning than to their orthography, supporting Li's (2015) finding that aptitude is more likely to be associated with learning in explicit conditions. The researchers also suggested that students drew on their aptitude to learn the word meanings and not word forms because the former is a more challenging task that requires more cognitive effort.

In our recent study (Muñoz et al., under review) we explored the effects of language learning aptitude on vocabulary learning measured by LLAMA D that tests the ability to recognise patterns in spoken language. The participants watched one captioned episode of a TV series (about 21 minutes) twice. Similarly to Suárez and Gesa (2019), it was found that the participants with higher aptitude scores performed better in the meaning recognition test.

Teng (2022) examined learning vocabulary from a 51-minute documentary with or without captions. The participants were undergraduate university students with intermediate English proficiency level. The results suggested that the LLAMA global score had a significant effect on receptive and productive vocabulary learning, as measured by four different tests on form and meaning recognition, and form and meaning recall. While the captioned group outperformed the uncaptioned group, the results for the effects of aptitude are not transparent. The author did not provide an interaction between the viewing group (captions or no captions) and aptitude scores, therefore it is still unclear whether language learning aptitude plays a major role in learning vocabulary through captioned or uncaptioned input.

While previous research has not examined the association between language learning aptitude and L2 grammar learning through the audio-visual input perspective, L2 grammar learning is the feature most frequently reported to be correlated with FL aptitude (Li, 2016), therefore it is warranted to explore the possible effects of FL aptitude on grammar learning through audio-visual input.

2.6 Linguistic input-related factors

Previous sections of this chapter have shown how learning from audio-visual input may be affected by learner-related factors such as proficiency, working memory capacity, and language learning aptitude. This section addresses the possibility of input-related mediating factors affecting this learning. Along with the different captioning modes discussed above in section 2.4, grammatical constructions learnability factors are included in the linguistic input-related factors. In particular, the grammatical constructions learnability factors such as frequency, construction type, and recency are discussed below.

2.6.1 Grammatical constructions learnability factors

As mentioned above, this study focuses on grammatical constructions as its target items. 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).

2.7 Learners' perspective of learning from audio-visual input

Previous sections of this chapter discussed the potential of language learning from audio-visual input, and the factors affecting it. In the final part of this chapter, the learners' perspective on learning from audio-visual input will be presented with the particular focus on feeling of learning and viewing preferences.

2.7.1 Feeling of learning from audio-visual input

While previous empirical research has pointed out the benefits of audio-visual input, and especially captioned videos (Montero Perez et al., 2013), it is also essential to take into consideration how students perceive the effectiveness of L2 audio-visual materials and captioning due to the possible effect these perceptions have on learning. Students' L2 learning beliefs represent an individual difference variable that influences both the process and outcome of language learning (Barcelos & Kalaja, 2011; Ellis, 2008). For instance, learner beliefs have been found to influence learner's attention to linguistic information

(Lai, 2019). That is, learners with positive attitudes towards the learning process are likely to achieve higher gains than those with negative attitudes (Ellis, 2008). Moreover, if there is a mismatch between the learner's and instructor's beliefs, little learning is likely to occur. Therefore, if L2 television viewers do not perceive this type of activity as useful for their language development, it will negatively affect their language uptake. Moreover, in a study on mental effort in learning (in the participants' L1) Salomon (1984) found that students perceived viewing TV easier than reading texts. The researcher suggested that this belief affected learning from the two types of input, with the texts leading to better outcomes as the participants regarded it as a more difficult activity, and therefore invested more mental effort into the task.

In this dissertation, learning beliefs are looked at as students' reflections about what language features they thought they had learned after a period of extensive exposure to L2 audio-visual input.

The audio-visual input studies discussing perceptions of learning from videos can be divided into two camps: research that explored overall feeling of learning from out of school exposure to audio-visual input (e.g. Dizon, 2021; Kusyk & Sockett, 2012) and intervention studies that included an exit questionnaire on feeling of learning after viewing (e.g. Montero Perez et al., 2014; Sydorenko, 2010; Winke et al., 2010). In general, out of school exposure questionnaire studies have found that language learners find original version audio-visual materials useful for their language development, especially for listening skills (Dizon, 2021) and vocabulary and expressions (e.g. Kusyk & Sockett, 2012).

The viewing intervention studies reported that participants perceived videos as being helpful for comprehension (Rodgers & Webb, 2017), vocabulary and expressions (Stewart

& Pertusa, 2004; Vanderplank, 1988), and speech segmentation (Winke et al., 2010). However, learners may not always be aware that learning is taking place, as was found in Sydorenko's study (2010) where the participants were asked whether they learned any L2 Russian vocabulary from viewing three video clips (2 to 3 minutes). While there were in fact significant vocabulary gains, many participants did not perceive any vocabulary learning from the treatment. The author suggested that the lack of feedback might have caused this incongruity and affected the perceptions of learning.

The majority of audio-visual studies focusing on learners' beliefs and attitudes towards audio-visual input included only a single viewing and did not look at perceptions of learning from extensive exposure. This is an important limitation, as Rodgers (2013) showed that pre-intermediate and intermediate participants' positive beliefs about content comprehension, listening ability, and vocabulary learning from videos increase over time as participants watch more episodes and become familiar with the TV series.

While Rogers' study provided valuable insights on attitudes and the change of perceptions over time, his study did not include different captioning modes. The presence or absence of captions, subtitles, or other forms of on-screen text that students experience during the viewing could play an essential role in the feeling of learning. Pujadas (2019) addressed this issue and compared the perceived usefulness of a longitudinal exposure to TV series between two groups of secondary school learners who viewed the episodes with either L1 subtitles (Spanish) or L2 captions (English). The results indicated that the captions group had a significantly higher feeling of learning than the subtitles group for vocabulary retention and learning.

Another longitudinal study (Mariotti, 2014) involved data collection on the perceived usefulness of audio-visual materials for language learning from several European countries and included extensive exposure to audio-visual input. The participants were asked to complete pre-exposure and post-exposure questionnaires and watch subtitled audio-visual materials with the audio in a target language (various European languages were included in the project), at least once in two weeks for nine months. The majority of participants reported an improvement in their listening skills, and almost 70% stated their intent to continue viewing subtitled materials. Unfortunately, the study referred to all on-screen text as subtitles and made no differentiation between different types of on-screen text (i.e. L1 subtitles and L2 captions), therefore it is impossible to tease apart the feeling of learning from two on-screen modes separately.

Avello (in preparation) explored how primary school children (ages 9–11) perceived learning from eleven captioned episodes of a cartoon TV series (approximately 120 minutes). The results of the post-viewing questionnaire indicated that the students felt they learnt vocabulary, phrases, and pronunciation the most, and grammar the least.

From the studies mentioned above, both those investigating single viewings and longitudinal research, it seems that learners are aware of learning taking place to some extent, with this awareness increasing with length of exposure to the audio-visual materials. However, none of these studies addressed proficiency differences, a factor that could be essential in the feeling of learning from audio-visual input (Teng, 2021: 37). For example, it was suggested that higher-proficiency learners have fewer negative beliefs about language learning than lower-proficiency students (Tang & Tian, 2015). In addition, the studies mostly focused on vocabulary and comprehension, and did not compare the feeling of

learning of various language features, such as words, expressions, grammar, and pronunciation. Although a number of studies (see Montero Perez, 2022 for a review) have provided evidence for the benefits of captioned over uncaptioned input for language gains, a comparison of the feeling of learning from different captioning modes is warranted as there is still a need to unveil whether learners perceive learning from various audio-visual input modes differently (e.g. L1 subtitles, L2 captions, and no textual support). This would also allow triangulation with the language gains observed in different studies with different types of on-screen text. In addition to investigating to what extent students feel like they are learning from audio-visual input, it is also important to think about students' attitudes towards viewing materials like these, particularly when combined with on-screen text. The next section discusses the importance of viewing preferences and how those are correlated with language learning from audio-visual input.

2.7.2 Viewing preferences and attitudes towards on-screen text modes

Viewing preferences is another warranted topic to explore as modes of viewing that are perceived more positively by learners are more likely to hold their interest and therefore enable greater amounts of language input.

The question of viewing preferences has been investigated through two different approaches: non-intervention questionnaire studies, and intervention studies with an exit survey. Several questionnaire studies from different parts of the world aimed to grasp participants' viewing habits (e.g. Dizon, 2021; Kusyk & Sockett, 2012; Muñoz, 2020). These reports show a wide variety in terms of on-screen text preferences and perceptions of usefulness of different captioning modes. For example, a comparison of Spanish and Danish

teenagers' out-of-school exposure revealed that national language viewing policies (i.e. dubbed or subtitled cultures) affect the viewing habits (Muñoz & Cadierno, 2021).

Other studies inquired about participants' viewing preferences after viewing interventions. For instance, intermediate-level French learners reported a preference for captions over no captions after viewing three clips of 16 minutes in total (Montero Perez et al., 2014). The participants claimed that the captioned mode was better for content and listening comprehension. As for the preference between L1 subtitles and L2 captions, the intermediate-level participants in Stewart and Pertusa (2004) watched two full-length Spanish films with Spanish L2 captions or English L1 subtitles. While the vocabulary gains were slightly higher in the captions group, there were differences in the attitudes towards two types of on-screen text. The participants who watched the films with L2 captions reported vocabulary learning and a clear preference for this type of input. Similarly, those who received the audio-visual input with L1 subtitles (that did not match the soundtrack), suggested that they could have benefitted more had they viewed the films with L2 captions instead, indicating their preference towards L2 captions rather than L1 subtitles. The L2 captions group also reported their intent to continue watching with L2 captions outside of the classroom for learning purposes.

Although the studies mentioned above demonstrated a positive attitude towards L2 captions, there is no consensus within language learners about the usefulness of captioning. For instance, Winke and colleagues (2010) asked their participants about their reactions towards watching three short (3-5 minutes) videos with or without captions. Overall, the learners perceived captions as a useful tool to chunk the speech sequence and as a helpful scaffold, though a few students were concerned about captions being more of a crutch and

distraction that prevented them from focusing on listening to the soundtrack and viewing the action in the clips. Similarly, in Teng (2021) some participants shared their concerns about the use of captioned video and suggested that the main issues were proficiency related, that is, due to having to split their attention between text and audio, they found that their proficiency level was not high enough to process either of the channels of input. Another disadvantage of viewing videos with captions mentioned by these participants was a worry about overreliance on on-screen text that they thought would never fade, supporting the previous suggestion in Winke et al. (2010) that captions may be perceived as "crutches".

Viewing preferences may also be affected by age and proficiency. Two studies based in different European regions (De Wilde et al., 2020 in Flanders; Muñoz & Cadierno, 2021 in Spain and Denmark) found that viewing with L1 subtitles was negatively associated with L2 knowledge, that is, participants with lower English proficiency tended to opt for L1 subtitles, while their more proficient peers preferred purely English input. It seems that learners who are older, who have higher language proficiency, and/or who are more experienced viewers, might opt for L2 captions instead of L1 subtitles (Muñoz, 2020).

The difficulty that lower proficiency learners might face while viewing captioned audio-visual input was reported in several studies. We saw above that Teng's (2021) participants felt their low proficiency reduced the value of audio-visual materials as a source of input due to the difficulty of focusing on multiple input channels. Another example is Taylor's study (2005), where it was reported that captions were distracting for lower proficiency level students who watched a 10-minute video in Spanish. In a recent study, De Riso (2021) found that 15 year-old Italian adolescents preferred watching original version English media with L1 subtitles, while their older 17 years old peers seemed to switch to

L2 English captions. Similarly, in Pujadas (2019) adolescent learners reported a shift in viewing preference before and after the extensive viewing intervention. The participants' preference for L1 subtitles decreased by 40%, and preference for L2 captions increased by 34.5%. At the same time, there was a small increase of 10% in students viewing without any on-screen text. Interestingly, the participants in the captions group reported their intention to continue viewing with L2 captions instead of L1 subtitles regardless of acknowledging that L2 captions were more demanding for them. Therefore, as a result of the viewing intervention, young participants saw a clear advantage of continuing viewing without any L1 support. In terms of proficiency differences, the lower-level students (about A1) reported a higher reliance on on-screen text while the participants of A2/B1 level considered on-screen text distracting. Overall, the self-reported data in Pujadas (2019) showed a shift from L1 subtitles towards L2 captions over time, especially for the more proficient participants, and those who were watching with L2 captions during the intervention.

While a number of studies have explored the viewing preferences and attitudes towards on-screen text, there is scarce research exploring changes in viewing habits of the same viewers (e.g. Pujadas, 2019). Mariotti (2014) suggested that regular exposure to audio-visual materials (once every two weeks for nine months) created a habit within the participants of viewing subtitled audio-visual materials. The questionnaire data also yielded that L1 subtitles were associated with leisure activities, while motivated learners preferred L2 captions. In other studies, it was also found that choosing one type of on-screen support over another might depend on students' confidence and familiarity with different viewing modes. For example, Vanderplank's (2019) participants reported less use of captions as they

became more familiar with the input. This finding was in line with Rodgers and Webb study (2017) where the more familiar the participants became with the TV series, the more confident they felt about viewing without captions.

2.7.3 Engagement with audio-visual input

Along with the viewing mode preferences, it is also important to take into account how learners engage with audio-visual input, as the ways L2 learners interact with this type of input may affect the learning outcomes. One of the variables that has been suggested to have an effect on learning from audio-visual input is familiarity with videos and frequency of exposure to L2 television, as Sockett and Kusyk (2015) found that frequent viewers were more likely to know frequently occurring chunks in a TV series (see above). In addition, a meaningful way in which L2 viewers may interact with audio-visual input is applying learning strategies- techniques to enhance the effectiveness of learning (Dörnyei, 2003). Kusyk and Sockett (2015) suggested that an interesting area of inquiry for future research would be analysing learners' viewing strategies to see what meaningful viewing activities (e.g. replaying the scenes, noting down unknown vocabulary, watching with L1 subtitles first and then with L2 captions) lead to greater gains. The authors suggested that it is unclear whether the students seize the opportunity to learn from L2 television, or are satisfied with it being a mere leisure pursuit. Learner's beliefs (see above) could be operationalised not only as what students think they are learning, but also what they consider as helpful learning strategies, as learner's beliefs regarding the language learning process are related to their use of strategies (Tang & Tian, 2015). Therefore, if learners, for instance, think that viewing L2 television is beneficial for their language learning, they may apply certain strategies to

enhance the learning experience. On the other hand, if the learners do not believe in the potential of original version materials and captions, they may merely watch the content without actively engaging with it.

This is a promising area of research as engagement with the audio-visual input, such as active viewing strategies, may facilitate learning from videos by making it a meaningful language activity rather than passive viewing. One of the few studies that discussed viewing strategies (Vanderplank, 2019) explored how intermediate and above L2 level learners interacted with captioned films outside of the classroom. The participants were asked to watch one film per week and document in their diaries the reflections of the experience. In particular, they were instructed to comment on their use of captions and what viewing strategies they applied. The author described the strategies that the participants used and identified two distinct groups of viewers in terms of strategies use: Minimal users who watched with infrequent pauses to check the meaning of unknown words or replayed difficult to understand scenes, and maximal users who utilised a variety of strategies such as noting down vocabulary and expressions and re-watching the same film multiple times with different captioning modes (e.g. the first time with captions and the second time without captions).

In another study, Alm (2021) asked several high-intermediate level learners of German to watch a season of a TV series of their choice at home and share their reflections about the experience in the classroom blog. The most reported strategy while viewing German TV series at home was watching the episodes twice. Repeated viewing seems to be a popular strategy not only with learners, but also teachers (Sergeeva, 2021), as language instructors report showing audio-visual materials to their students several times: First

without captions to access overall comprehension, and then with captions to pay attention to specific language features appearing in the video. Another strategy reported in Alm (2021) was avoidance of looking up vocabulary while viewing; the participants would rather enjoy the TV series and grasp the meanings from the context cues, than pause to translate an unknown word. This is in line with Fievez et al. (2021) where the participants did not always make use of the glossed captions, partly because the captions were too fast to access the meaning, but also because they would rather not translate every unknown word and watch without interruptions.

It was suggested that proficiency is one of the factors that may affect learners' choice of strategy use (Oxford, 1989), and that students with higher proficiency levels employ more language learning strategies (Tang & Tian, 2015). The abovementioned studies on viewing strategies explored engagement with audio-visual materials of intermediate level participants and above, and there have been no studies that have investigated the viewing strategies of lower proficiency students, or that compared different proficiency groups in term of their viewing strategies.

Finally, given that attitudes towards captioned audio-visual input and its use seem to change over time (Vanderplank 2019), and that learner's beliefs about how languages are learnt are not stable (Ellis, 2008), it is reasonable to assume that viewers are likely to switch from one viewing strategy to another (e.g. from a minimal user to a maximal user and vice versa). Therefore, it would be crucial to observe what factors affect this change (e.g. familiarity with the input, change of captioning mode.

Chapter 3

Rationale and research questions

The review of the literature above has shown mixed and inconclusive results for grammar learning through audio-visual input, therefore more research is needed to account for differences across and within grammar studies. This is especially true in the under-represented perspective of construction grammar, and whether L2 learners can find the constructions' patterns through prolonged exposure to audio-visual material without focus on form teaching.

As for learner-related factors, learner proficiency appears as an important variable in some studies and as insignificant in others, and the uncertainty of those findings is increased because the studies have mostly focused on intermediate and high-intermediate proficiency learners. Regarding working memory capacity, the scarce results do not yet provide strong confirmation of the central role attributed to it by information theories of multimedia learning. Furthermore, more research is undoubtedly needed that explores the role of language learning aptitude, or specific aptitude components, in grammar learning through audio-visual input.

Regarding input-related factors, mixed results have been found for different captioning modes (i.e., captions, textually enhanced captions, and no captions), and no grammar studies have yet looked at prolonged exposure to audio-visual input supported by various types of on-screen text. Moreover, no audio-visual input study so far has looked at

the learnability factors of grammatical constructions (i.e., construction type, frequency, and recency).

Several audio-visual input studies have aimed to explore learners' feeling of learning, viewing preferences and strategies through post-viewing questionnaires or interviews. One thing to note about many of these studies is the preponderance of qualitative data often from small populations or analysis that do not venture beyond descriptions. This leads to conclusions that might not be generalizable and warrants a study that also integrates inferential statistics methods. Moreover, to the author's knowledge there has only been one study that documented a longitudinal change of viewing preferences (Pujadas, 2019).

The aim of this dissertation is to focus on the emerging gaps in the area of learning grammar through audio-visual input. In order to do so, three related studies were conducted to address the abovementioned research gaps.

The aim of Study 1 is to explore the extent to which L2 grammar and specifically constructions can be learnt from watching captioned/uncaptioned full-length episodes of a TV series, and the role that three learner-related factors: proficiency, WM capacity, and language learning aptitude may play. Study 1 sets out to answer the following research questions:

Research question 1: To what extent can L2 constructions be learnt through extended exposure to TV series?

Research question 2: What are the roles and relative importance of captions, proficiency, WM capacity, and language aptitude in learning L2 constructions?

Study 2 aims to examine the potential benefits of different types of captioning (captions, TE captions, and no captions) on grammar learning and the effects of three factors that may be determinants of construction learnability: construction type, frequency, and recency. The following research questions are addressed in Study 2:

Research Question 3: To what extent does L2 construction learning from audio-visual input depend on caption support, construction type, frequency, and recency?

Research Question 4: In what ways do these factors interact in L2 construction learning from audio-visual input?

The aim of Study 3 is fourfold: explore students' feeling of learning from extensive exposure to full-version TV series, explore the role of proficiency and intervention captioning mode (with captions, without captions, and with enhanced captions) on that feeling of learning, see if there is a preference change in their use of captioning mode and whether this is also affected by students' proficiency and intervention viewing mode. Finally, the study also attempts to discover the viewers' learning strategies while watching audio-visual materials, and to ascertain whether those strategies are associated with proficiency level. Study 3 aims to address the following research questions:

Research question 5: To what extent is students' feeling of learning from audio-visual input affected by intervention viewing mode and proficiency?

Research question 6: To what extent do students' viewing preferences change over time, and does this change depend on proficiency and/or intervention viewing mode?

Research question 7: What is the use of viewing strategies by different proficiency groups before and after the intervention?

Chapter 4

Methodology

This chapter presents the methodology of the three studies in this doctoral dissertation. It provides an overview of the characteristics of the participants, audio-visual materials, target grammatical constructions, instruments, and procedures which were common to all three studies. The specific methodologies that applied to each study are presented later, in the study-related chapters below.

In order to address the research questions described above, an experimental design was chosen, with an ecologically sound in-class intervention. The participants were shown original version audio-visual material with an experimental variable of 'type of on-screen text'. Through the use of a pre/post/delayed post-test design, quantitative analysis was performed to answer research questions 1, 2, 3, and 4 regarding learning from audio-visual input. Additionally, questionnaires were used to answer research questions 5, 6, and 7, and triangulate the data.

4.1 Participants

The participants for this study were Catalan/Spanish bilinguals, their ages varied from 17 to 32 years old (M = 19.32, SD = 2.4). They were recruited at the Faculty of Information and Audiovisual Media of the University of Barcelona where they were pursuing undergraduate degrees in Audiovisual Communication. The participants were attending a compulsory course on Oral and Written Communication in English. This course is unstreamed and always includes students from a variety of proficiency

backgrounds, from A1 to C2. The content of the course was vocabulary based and related to students' majors (advertising, cinema, or marketing), the curriculum did not cover grammar practice. The intervention was integrated as a part of the language course; all procedures were carried out during class time, and students were allocated course credits for participating in the study.

All participants had the same language teacher and were not informed about the nature of the experiment beforehand. The course lecturer is an experienced scholar familiar with classroom interventions and data collection methods for audio-visual input research. The lecturer was explicitly asked by the researcher not to provide any instruction of the target grammatical constructions.

Four intact classes, with a total of 175 students participated in the study. After a thorough examination of the students' previous exposure to the TV series, attendance of the viewing sessions, and test (grammar and cognitive) completion, 34 participants were excluded from the dissertation's analyses, leaving 141 participants.[2]

According to the pre-course questionnaire (see below), the participants were familiar with viewing media in English: 94% of students indicated that they watched films and TV series in English on a weekly basis, 18% watched English movies and TV series daily, 19% watched movies or TV series 4-6 times a week, 24% viewed audio-visual materials between 1-3 times a week, and 33% watched videos once a week. When asked about their use of on-screen text while watching original version materials, 56% reported viewing with L1 subtitles, 78% viewing with English captions, and 27%

[2] Note that the total number of participants varies across the three studies of this doctoral dissertation as they, for example, missed the cognitive test session (Study 1) or belonged to a lower proficiency group (Study 2), therefore some participants were included in only one study, and others in two or three.

viewing without any text support (the percentages do not add to 100 as the participants could choose various options, and their viewing preferences were generally not limited to just one viewing mode).

The intact classes were randomly assigned to a viewing condition. The intervention took place in the autumn semesters of 2018 and 2019. The intervention for the captions (n=39) and the no captions (n=30) groups took place in autumn 2018. The data for the textually enhanced (n=39) and the captions (-) (n=33) groups were collected in autumn 2019. The data were collected over two years due to the availability of the intact classes. The lecturer who helped with the data collection taught the course in autumn semester and had two parallel groups at a time. This allowed us to analyse the results of the captions and no captions groups first (see Chapter 5), then during the second round of the data collection, to address some issues that appeared as promising areas for further research. In particular, the textually enhanced captions group with the target constructions highlighted in yellow and bold, and the captions (-) group who did not complete the immediate post-tests (see below) was added to provide evidence of the potential effect of the immediate post-tests on learning.

The participants' English proficiency scores ranged from 90 to 183 according to the Oxford Placement Test (OPT) results (this is equivalent to CEFR levels from A1 to C2, with a mean of B2). Their mean vocabulary size was 3373 words (as measured by X_Lex, Meara & Milton, 2003), which falls within the band of upper-intermediate level (B2) (Milton & Alexiou, 2009). Table 1 presents the ages, proficiency levels and vocabulary sizes by intervention group.

Table 1. Participants' age, proficiency, and vocabulary size

Group		Age			Proficiency (OPT) (max: 200)			Vocabulary size (max: 5000)		
	n	Mean (SD)	Min	Max	Mean (SD)	Min	Max	Mean (SD)	Min	Max
Captions	39	19 (2.79)	17	32	140 (21.20)	92	178	3377 (763)	1600	5000
No captions	30	19 (2.55)	17	30	137 (24.54)	94	183	3121 (693)	1900	4250
TE captions	39	18 (1.40)	18	23	132 (17.14)	90	168	3202 (620)	2050	4650
Captions (-)	33	19 (1.21)	18	23	139 (17.14)	99	164	3207 (604)	2000	4250
All participants	141	19 (2.11)	17	32	137 (20.04)	90	183	3236 (674)	1600	5000

4.2. Materials

4.2.1 Video selection and characteristics

After a careful analysis of several TV series, *The Good Place* (Schur, 2016) was chosen for the purposes of this doctoral dissertation. The selection included the following criteria:

1) The TV series must not have been broadcast on Spanish central TV.
2) The TV series should be relatively new and unfamiliar to the participants.
3) The TV series should include a variety of grammatical constructions.
4) The TV series should be appropriate for a classroom viewing (i.e., minimal violence, cursing, nudity).

The Good Place is an American fantasy comedy first released in 2016 with a plot revolving around a young woman, Eleanor Shellstrop, who dies and is assigned to spend

her afterlife in 'the Good Place'. The Good Place was specifically created for outstanding individuals who deserved to spend the eternity enjoying themselves after dedicating their lives to helping other people. Although at first Eleanor thinks that she belongs in the Good Place, she soon realizes that she was put there by mistake and that she must hide her relatively immoral past (compared to the rest of the inhabitants of the Good Place). She finds refuge in a university professor of ethics and moral philosophy who starts teaching her how to become a better person. No version of *The Good Place* (neither original version nor dubbed) had been shown on Spanish Central TV at the time of the intervention (nor has it been up until the publication of this thesis).

In order to control for any previous exposure to the TV series through *Netflix* or other means of streaming, the pre-course questionnaire (see Appendix A) included a question on whether the participants had watched *The Good Place* before. Only one participant had watched the show before the intervention and therefore their scores were not included in the analysis.

To identify whether the TV series had a variety of grammatical constructions, the scripts of the first seasons were run through a Python script using the Natural Language Toolkit (Bird et al., 2009) that identified the most frequently occurring n-grams of various lengths from 2 to 5 words. A detailed description of the target constructions selection from *The Good Place* is provided below.

Finally, the content of the show was appropriate for classroom use as the in-universe rules of the Good Place precluded anything bad from happening. For example when a character tries to curse, she is unable to, and instead is forced to say the word 'fork'. On another occasion, when a dumpster falls on a pair of characters, they immediately jump up through the dumpster, having suffered no harm. The show

generally has a light, comedic tone that made it both enjoyable and particularly suitable for use in the classroom.

To ensure that the participants could follow the TV series, the lexical profile of the episodes was analysed through *Range for texts* (Cobb, n.d.) in the *Lextutor* software (https://www.lextutor.ca/cgi-bin/range/texts/). This analysis was performed to establish whether the participants had the necessary vocabulary level to understand the dialogue of the show. As can be seen from Table 2, 89% of words fell within the first 1000 words band, and 95% of the vocabulary used in the TV series was within the 3000 words band. The vocabulary demands of the target TV series were therefore suitable for our viewers to follow the content, as 95% coverage is needed for sufficient comprehension of a text (Laufer, 1989) or a TV programme (Webb & Rodgers, 2009).

4.2.2 Selection of the target constructions

This study situates itself in the constructionist perspective of grammar that states that learning a language is learning its constructions (Ellis et al., 2016). Those constructions can differ in their length, frequency, and complexity. The list of the target grammatical constructions was formed through a careful analysis of the TV series transcripts. Given the importance of frequency for grammatical constructions learning (see above), it was decided that the target constructions should appear in the TV series at least three times. First, the transcripts of the ten episodes were run through the n-gram Python script (Bird, et al., 2009) to identify frequently occurring 2-grams, 3-grams, 4-grams, and 5-grams. The analysis did not yield any frequent results for n-grams containing more than five words, the script is provided in Appendix B. Second, all the transcripts were analysed manually to ensure that the natural language processing script had not missed any frequently occurring constructions, or misconnected any phrases.

For instance, the irregular plural construction was not identified through the n-grams script because it consists of one word. Third, the constructions were lemmatized and grouped into categories. For example, the n-grams that included the conjugated verb *to be+ 3rd participle* were grouped as *passive.* Finally, the tentative list of constructions was shown to the class teacher who reflected on the complexity of the chosen constructions and confirmed that the variety of the target forms was suitable for the intact classes. As the students were not assigned to the English class by language level, and there were students from A1 to C2 levels in the same classroom, it was important to select the constructions that would be suitable for a variety of levels.

Table 2. Vocabulary coverage of The Good Place chosen episodes

Episode		K1		K2		K3		Cumulative range within K3	
	N of words	n	%	n	%	n	%	n	%
Episode 1	3189	2875	90.15	148	4.64	51	1.63	3074	96.42
Episode 2	3062	2758	90.07	125	4.08	57	1.89	2940	96.04
Episode 3	3287	2970	90.36	145	4.41	47	1.43	3162	96.20
Episode 4	2979	2629	88.25	140	4.73	59	1.98	2828	94.96
Episode 5	3264	2923	89.55	169	4.66	73	2.24	3495	96.45
Episode 6	2991	2710	90.61	110	3.68	53	1.77	2873	96.06
Episode 7	3128	2811	89.87	161	5.15	59	1.89	3031	96.91
Episode 8	2865	2551	89.04	132	4.61	59	2.06	2742	95.71
Episode 9	3379	3040	89.97	128	3.79	46	1.36	3214	95.12
Episode 10	3365	3015	89.60	163	4.84	53	1.58	3231	96.02
Total	31509	28282	89.76	1417	4.50	557	1.77	30245	95.99

4.2.3 Target constructions

Following this selection process, a total of 27 constructions were targeted on the basis of their frequency of occurrence: only constructions that appeared at least three times in the ten episodes (227 minutes) were included in the study. This dissertation adapts Fried's (2015) classification of constructions and distinguishes between fully-filled constructions (fixed multiword units with no variation in the input, e.g. *say no more*), partially-filled constructions (with at least one variable slot, e.g. *the Xer, the Yer*), and fully-schematic constructions (e.g. future in the past). This classification was used as, compared to other classifications (see Chapter 2), it represented the target constructions the best. Given that constructions are learnt through the specific input which learners are exposed to (Goldberg & Casenhier, 2008), the target constructions were allocated to these three groups depending on the specific language exemplars present in the input material: fully-filled (6), partially-filled (11), and fully-schematic (10). See Table 3 for a list of target constructions, with real examples taken from *The Good Place* script (with the target constructions in italics), their categorisation, and frequency of occurrence in the TV series and in the TV Corpus (Davies, 2019). The TV Corpus is a part of The Corpus of Contemporary American English, and it contains informal language from TV shows released between 1950 and 2018. The target constructions can be claimed to be ecologically valid and representative of other English media sources because their frequency of occurrence in *The Good Place* TV series and in The TV Corpus are strongly correlated ($r = .877$, $p < .001$).

Regarding the frequency of constructions encounters, the constructions were presented as they occurred throughout the episodes, mixed together naturally. Each episode contained between 17 and 40 total target constructions, and 7 to 16 different

target construction types. The distribution of the target constructions throughout the 10 episodes is presented in Table 4.

4.3 Instruments

4.3.1 Pre-test, immediate post-test, and delayed post-test

The testing materials to evaluate target constructions knowledge included productive grammar exercises, such as sentence transformation, fill-the-gap, and complete the gap with the correct form of a given word (the examples of the test items are presented in Figure 1 below). These types of exercises were chosen due to students' familiarity with the format (from language coursebooks). The target constructions were tested with different types of test items (see Table 5) due to the different characteristics of the constructions. For example, the construction *to be allowed to* could not be tested with the "complete the gap with the correct form of a given word" without giving away too many details, therefore a "sentence transformation" exercise was chosen instead. Similarly, the irregular plural of the nouns could not have been tested with the "sentence transformation" test, so "complete the gap with the correct form of a given word" exercise was used.

A total of 108 test items were created. Each test included two instances of the target constructions in order to more accurately assess the participants' knowledge, giving them two opportunities to demonstrate that they knew the constructions.

Table 3. A list of target constructions with frequency of occurrence

Construction type	Construction form and test type	Examples from The Good Place	Frequency	
			The Good Place	TV Corpus
Fully-filled	do for a living	What did you *do for a living*?	3	1214
	let you down	I won't *let you down*	3	5456
	N[irregular plural]	There are *shrimp* flying around	5	3352
	big deal	No *big deal*	6	13663
	say no more	*Say no more*	3	754
	figure out	To *figure out* what's going wrong	15	29301
Partially-filled	to be[tense] allowed to V	I'*m not allowed to* tell you about	3	10158
	would rather V	I'*d rather* not *let* people see it	3	1359
	break[tense] DET promise	You *broke your promise*	3	292
	the Xer the Yer	*The more* you practice, *the more* you improve	4	188
	used to V	I used to just throw them in the sink	7	23329
	PRON just want[tense] to	*I just want to* be an academic	9	36160
	let's V, shall we?	So let's chat, *shall we*?	3	4984
	why don't PRON	*Why don't you* go ahead?	12	51952
	to be[tense] supposed to V	You *were supposed to* be there	18	83675
	subj belong[tense] here	You don't *belong here*	18	1501
	let's V	*Let's* move on	54	246479
Fully-schematic	passive present continuous (subj aux VP)	Her memory's still *being rebooted*	3	22897
	future continuous (subj aux V-ing)	Later this evening, we *will be enjoying*	3	6475
	subjunctive (subj V that PRO V)	You *wish* that you *were* related	4	2121
	V[negative] either	You're not supposed to be here *either*?	5	7520
	passive present perfect (subj aux VP)	It *has been proven*	3	13588
	reported speech (reporting V (that) V)	Tahani *said* that you *helped* Michael	6	4371
	catenative V obj infinitive (sub V PRO to V)	You *need me to* lie	17	128926
	catenative V obj bare infinitive (let PRO V)	Should I *let her stay*?	20	126232
	future in the past (subj V[past] V)	I thought transition *would* be easier	14	3644
	emphasis (do[tense] V)	That *does* sound like me	25	95969

Table 4. Distribution of the target constructions in the episodes

Construction type	Construction	E1	E2	E3	E4	E5	E6	E7	E8	E9	E10	Total
Fully-filled	do for a living	1			2							3
	let you down		2				1					3
	N[irregular plural]	3		1				1				5
	big deal					6						6
	say no more				1		2					3
	figure out		3	2	2		2		3	1	2	15
Partially-filled	to be[tense] allowed to V	1					2					3
	would rather V					2				1		3
	break[tense] DET promise						3					3
	the Xer the Yer							2	1	1		4
	used to V			1		2			2	2		7
	PRON just want[tense] to			2	2			2	1		2	9
	let's V, shall we?	1		1							1	3
	why don't PRON			2	1	5				3	1	12
	to be[tense] supposed to V	3		3	3		2	3	2	1	1	18
	subj belong[tense] here	1		8	2				2	3	2	18
	let's V	5	5	5	4	2	7	2	3	15	6	54
Fully-schematic	passive present continuous (subj aux VP)	1						2				3
	future continuous (subj aux V-ing)				1		2					3
	subjunctive (subj V that PRO V)								2	1	1	4
	V[negative] either	1		1	1	1				1		5
	passive present perfect (subj aux VP)	2		1								3
	reported speech (reporting V (that) V)			1		2	3					6
	catenative V obj infinitive (sub V PRO to V)	1	2	3	1	2	2	1	2	2	1	17
	catenative V obj bare infinitive (let PRO V)	3	1	3		1		1	5	4	2	20
	future in the past (subj V[past] V)		3	1	3	1	2		1	2	1	14
	emphasis (do[tense] V)	3	1	4			1	3	5	3	5	25
	Total	26	17	39	23	24	29	17	29	40	25	269

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 target constructions. 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 none of the test items presented the target constructions to students; it was the students who had to analyse the prompts and produce the constructions themselves.

The pre- and delayed post-tests consisted of 54 test items (two test items per target construction) and both pre- and delayed post-tests contained the same test items in a randomized order. Cronbach's alpha showed that both the pre-test ($\alpha = 0.879$) and the delayed post-test ($\alpha = 0.835$) reached an acceptable level of internal consistency.

I. Sentence transformation exercise:

*Complete each sentence with **two to five words**, including the word in bold*

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 exercise:

Complete the sentences using a form of the words in brackets

_____ (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:

Let's go to the theater, _____ we?

Figure 1. Examples of the test items

An immediate post-test was administered in the second class of each week (see Figure 2 in the Procedure section below). The aim of this test was to contrast immediate construction learning with cumulative learning that was measured in the delayed post-test. Two post-tests were included in the design to analyse the effects of recency on grammatical constructions learning (see Chapters 2 and 7). The immediate post-test included test items on the five to seven target constructions that had appeared with the highest frequency in the preceding two episodes in that week. Those tests had the same format as the pre-/delayed post- tests (productive grammar exercises), but had different test items to avoid practice for the delayed post-test. In total, the five partial immediate post-tests consisted of 54 test items with two items per construction, and every construction was tested once throughout the partial immediate post-tests (see Table 5 for the constructions' week of testing and test type). The Cronbach's alpha for the immediate post-test reached an acceptable reliability level ($\alpha = 0.795$). The pre-test, and delayed post-test are provided in Appendix C, and the delayed post-test in Appendix D.

As regards scoring, one point was assigned for a correct answer per question in the pre-test, immediate post-test, and delayed post-test. 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.

Table 5. Immediate posttest. Week of constructions' testing and test items

	Constructions tested in the partial immediate post-tests
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

4.3.2 Individual differences tests

Three individual differences tests were also run in order to control for any differences between groups, and to investigate whether these variables had any impact on learning from audio-visual materials (research question 2). These variables were English language proficiency, working memory (WM) capacity, and language learning aptitude.

Participants' proficiency was measured by means of the pen-and-paper version of the Oxford Placement Test (OPT) (Allan, 2004). 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 score of 90–104 refers to A1, a basic level, extremely limited user; 105–119 to A2, elementary level, limited user; 120–134 to B1, lower intermediate level, modest user; 135–149 to B2, upper intermediate level, competent user; 150–169 to C1, proficient level, advanced user; and finally 170-189 score refers to C2, highly proficient level, very advanced user.

Along with the proficiency test, the X_lex test was used to measure participants' vocabulary size (Meara & Milton, 2003). This test consists of 120 items and measures the knowledge of the most frequent 5000 words in English[3].

The WM capacity test was computer-delivered and the results were scored automatically. It was measured by means of an L1 reading span task (Gilabert & Muñoz, 2010) via the E-Prime 2.0 software. The participants could choose whether they wanted to complete the task in Catalan or Spanish. The participants performed the WM test in their L1, as previous research has suggested that performing the test in the participants' L2 may affect the results due to limits in L2 knowledge (Sagarra, 2017). The task has the following procedure: The test takers have to read a string of letters (from two to nine) and try to remember those letters, then the screen changes, and they have to read a sentence and decide whether it makes sense or not (half of the sentences make sense, and the other half do not). Finally, the participants are asked to recall the letters in the order they appeared. Then the cycle repeats and gets more and more complex (i.e., the

[3] At the moment of the data collection the X_Lex test was available on Lognostics software (Meara & Milton, 2003); however, as of February 2022 this test is no longer supported, and has been superseded by the Yes/No test (Meara & Miralpeix, 2017).

participants are presented with longer strings of letters). Accuracy, order, and reaction times are considered in the final score.

Language learning aptitude was tested with the use of the language-independent LLAMA aptitude test (Meara, 2005). This test comprises four sub-tests (LLAMA B, LLAMA D, LLAMA E, and LLAMA F) based on an unfamiliar language. The LLAMA tests were also computer-based and the results were scored automatically by the software. LLAMA B is a test of associative memory where the participants are presented with twenty pictures, each with a short, unfamiliar name that can be displayed when they are clicked on, and the participant must learn as many names of the objects as possible in two minutes. LLAMA D predicts the ability to recognize patterns in spoken language. In the learning phase, the test takers listen to a sequence of exotic words from the unfamiliar language, and in the test phase they have to decide whether the word they hear is from the learning phase or a new one. LLAMA E tests sound-symbol association, the participants have to learn the sounds and symbols of a new language. Finally, LLAMA F measures the analytical ability to infer and learn a grammatical structure of the unknown language. The test takers are presented with images and sentences representing grammatical rules that they then have to apply later in the test phase. It has been suggested that this battery of tests assesses two different aptitudes: explicit learning aptitude (LLAMA B, E, and F), and implicit learning aptitude (LLAMA D) (Granena, 2013).

4.3.3 Questionnaires

An adaption of the GRAL research group questionnaire on out-of-school exposure to English (e.g. Muñoz, 2020) was used to develop two sets of questionnaires on exposure to audio-visual input outside of the classroom. The pre-course questionnaire (Appendix A) consisted of thirteen questions that addressed the participants' engagement with English media and on-screen text (see the Participants section above). Importantly, the questionnaire included a question on previous viewing of *The Good Place* TV series. The post-viewing questionnaire (Appendix E) included the same questions as the pre-course survey, along with questions on feeling of learning from *The Good Place*.

4.3.4 Viewing intervention reflections

After the viewing intervention the participants were also asked to write a composition reflecting on their classroom experience of watching the TV series. It was important to give participants voice to reflect on the experience to grasp their beliefs about learning from the TV series as learner's perceptions may affect the learning outcome (Ellis, 2008). In addition, the participants' reflections could provide a triangulation of the questionnaires responses (see above). Finally, this composition embedded the viewing intervention into the classroom assessment. The participants were asked to share their thoughts about using the TV series in the classroom, express their preference towards viewing with or without captions, and share whether they thought they would learn anything from this experience (see Appendix F).

4.3.5 Post-viewing activities

In order to integrate the viewing into the classroom, the participants were asked to complete several post-viewing activities after viewing each episode. These activities included a recognition task and a comprehension task. The purpose of these exercises was to keep students' attention focused on the language and content of the videos. The recognition task asked students to decide whether they had seen/heard or heard (depending on the condition) phrases presented in the episode they just watched. Half of the phrases came from the recently watched episode and the rest were distracters that did not appear in the TV series, but none were target constructions. The content comprehension task was a multiple-choice task that was not focused on information related to the target constructions. The post-viewing activities per each episode are presented in Appendix G. The results of these activities are not analysed in this dissertation as this falls beyond the scope of the present studies.

4.6 Procedure

The procedure is summarized in Figure 2. The intervention took place twice a week (90 minutes per class) during regular class time over a period of eight consecutive weeks in the first semester of the academic year. The students were not notified beforehand about the upcoming tests the following classes. Students were given class credits for watching the episodes and completing the tests.

The first two weeks students completed both parts of the OPT (60 minutes), the pre-test (40 minutes), vocabulary size test (10 minutes), WM test (20 minutes), language learning aptitude test (40 minutes), and the pre-course questionnaire (10 minutes). The

tests and tasks were administered by the class teacher, except for the cognitive tasks, for which several researchers assisted with the data collection.

During each of the following five weeks, the participants watched two full-length episodes of the TV series on two different days (Tuesdays and Thursdays). The TV series DVD was used to show the episodes on the classroom screen, the English captions option was included in the DVD settings. Every week followed the same protocol: On Tuesdays students watched one episode (about 22 minutes) and completed post-viewing activities (10 minutes); on Thursdays students watched the next episode and completed post-viewing activities (10 minutes) together with the immediate post-test (10 minutes). The captions (-) group followed the same protocol but did not complete the immediate post-test and continued with the class instead. This allowed us to control for a potential testing effect of the immediate post-test on the delayed post-test.

Students completed the delayed post-test and post-viewing questionnaire in the last week of the intervention, five days after watching the last episode.

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 (the procedure of textual enhancement is described in Chapter 6), and the no captions group watched the episodes without captions.

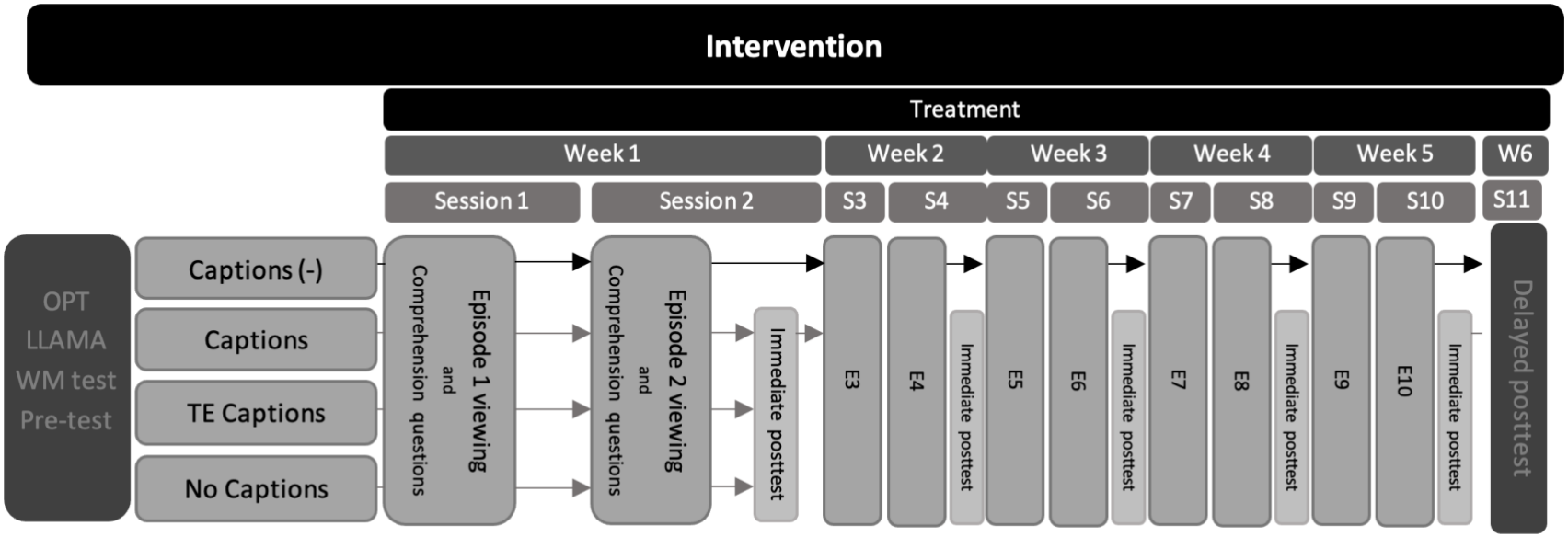

Figure 2. Pedagogical intervention

Chapter 5

Learning L2 constructions from captioned audio-visual exposure: The effect of learner-related factors

5.1 Introduction

This chapter presents a study about learning of English grammatical constructions from original version TV series is analysed. Although the research is moving in the direction towards analysing grammar learning through audio-visual input, no studies so far have looked at grammar learning through prolonged viewing of original version materials. Therefore, it is warranted to understand whether exposure to such materials as TV series in the target language may lead not only to significant vocabulary gains, but also grammar learning.

Another area of inquiry in this study is the effects of learner-related factors. First, proficiency has been claimed to have a mediating role in learning from captioned audio-visual input (e.g. Danan, 2004), however, the research has yielded mixed results on the matter. Second, the role played by cognitive individual differences, such as working memory capacity and language learning aptitude also remains mostly unexplored in audio-visual input research, regardless of the language feature.

The study presented in this chapter uses an original version TV series to examine the effects of captioned and uncaptioned audio-visual input on learning L2 grammar. The

grammar focus of this study is the acquisition of a set of grammatical constructions (see Chapter 6 for a study that distinguishes different types of constructions: fully-filled, partially filled, and fully-schematic). The study also investigates the role of three learner-related factors: proficiency level, WM capacity, and language learning aptitude. The following research questions are addressed in this study:

Research question 1: To what extent can L2 constructions be learnt through extended exposure to TV series?

Research question 2: What are the roles and relative importance of captions, proficiency, WM capacity, and language aptitude in learning L2 constructions?

5.2. Methodology

5.2.1 Participants

The participants from the first year of the data collection (autumn 2018) were included in this study. A total of 69 participants' results from the captions (n=39) and no captions (n=30) groups were analysed. Their English proficiency scores ranged from 92 to 183 (M= 139.55, SD= 22.32) according to the Oxford Placement Test results (from A1 to C2, with a mean of B2).

Both groups watched the same audio-visual material and completed identical tests, the only difference was that one group watched the TV series episodes with captions and the other without captions. The description of the materials and procedure is provided in Chapter 4.

5.2.2 Scoring and data analysis

The present study analysed the pre-test and the delayed post-test results of the intervention. The immediate post-test results, that is, the results of each weekly post-test, were not incorporated in the analyses of this study because the aim is to analyse the overall learning from the intervention. The delayed post-test is further referred to as *post-test* in this chapter.

One point was assigned for a right answer per item; a correct answer included all parts of a construction in the correct form but students were not penalized for spelling mistakes. The data analysis was run in the SPSS version 25 statistical package. First, several independent samples t-tests were run to ensure that the two groups were comparable before the intervention. Next, the Wilcoxon Signed Rank test was used to answer the first research question. Finally, a series of General Linear Models (GLMs) were run to answer the second research question. The GLMs were chosen to address the second research question as this analysis allows to include several fixed factors, covariates, and interactions in the same model. It was important to include the covariates to control for the previous knowledge in the pre-test.

5.3 Results

The descriptive statistics of the variables of the two groups in the study (captions and no captions) are presented in Table 6 below. To test whether there were any differences between the captions and no captions groups before the intervention, four independent samples t-tests were conducted. The results showed that both groups were comparable as there were no significant differences between their pre-test scores (p = .986), proficiency scores (p = .372), working memory scores (p = .428), and aptitude scores (p = .080).

5.3.1 Research question 1: Learning L2 construction through audio-visual input

The first research question addressed the effect of audio-visual input on the learning of L2 constructions. As can be seen from Table 6, the participants (in both groups) showed evidence of learning through their improved performance in the post-test (regardless of condition). The initial set of analyses examined the impact of watching the episodes in the original version by comparing the pre- and post-test scores of both groups together. The data for the pre-test scores were not normally distributed (p < .001) and, for this reason, the Wilcoxon Signed Rank test was used. The test revealed a significantly higher performance in the post-treatment test with a medium effect size (Plonsky & Oswald, 2014) (Z= 6.234, p < .001, r = .755) indicating that overall the participants improved their knowledge of the target L2 constructions after the intervention. The percentage of the learnt target constructions is presented in Table 7.

Table 6. Descriptive statistics per experimental group

	Pre-test score (max: 54			Post-test score (max: 54)			Proficiency (max: 200)			WM (max: 81)			Aptitude (max: 400)		
	Mean (SD)	Min	Max	Mean (SD)	Min	Max	Mean (SD)	Min	Max	Mean (SD)	Min	Max	Mean (SD)	Min	Max
Captions (n = 39)	23.62 (11.52)	5	45	32.13 (10.99)	12	49	140.85 (21.20)	92	178	35.10 (13.20)	9	62	233.08 (37.28)	165	320
No captions (n = 30)	24.07 (11.17)	2	45	30.07 (13.18)	6	51	137.87 (23.96)	94	183	33.83 (13.54)	11	62	212.17 (60.09)	60	330
All participants (n= 69)	23.81 (11.55)	2	45	31.23 (11.94)	6	51	139.55 (22.32)	92	183	34.55 (13.27)	9	62	223.99 (49.25)	60	330

Table 7. Percentage of learning of the target constructions

Construction type	Target construction	Percentage (%) of possible learning	
		Captions	No captions
Fully-filled	do for a living	25.20	15.10
	let you down	43.40	36.30
	N[irregular plural]	27.70	13.00
	big deal	24.40	36.60
	say no more	50.00	28.50
	figure out	2.90	0.00
Partially-filled	to be[tense] allowed to V	69.60	56.20
	would rather V	11.10	12.10
	break[tense] DET promise	46.60	50.00
	the Xer the Yer	22.60	15.00
	used to V	45.40	41.10
	PRON just want[tense] to	60.00	44.40
	let's V, shall we?	50.00	36.80
	why don't PRON	24.10	32.30
	to be[tense] supposed to V	70.00	54.50
	subj belong[tense] here	59.60	50.00
	let's V	59.20	61.90
Fully-schematic	passive present continuous (subj aux VP)	36.80	41.60
	future continuous (subj aux V-ing)	60.00	26.60
	subjunctive (subj V that PRO V)	30.70	25.00
	V[negative] either	24.50	24.20
	passive present perfect (subj aux VP)	44.40	41.60
	reported speech (reporting V (that) V)	29.70	26.00
	catenative V obj infinitive (sub V PRO to V)	40.00	30.00
	catenative V obj bare infinitive (let PRO V)	33.30	40.00
	future in the past (subj V[past] V)	28.00	12.50
	emphasis (do[tense] V)	20.30	26.00

5.3.2 Research question 2: Factors explaining learning

To address the second research question, General Linear Models (GLMs) were run in order to explore the influence of group (captions vs no captions), proficiency, WM capacity, and aptitude on the participants' learning of constructions. The correlations between WM scores were low and non-significant with both total LLAMA scores (r = .032, p = .792), and LLAMA subtests separately (p >.05). Therefore, these two constructs were analyzed separately as independent factors.

Data exploration showed that the variables proficiency and aptitude were not linearly distributed. The variable proficiency was recoded into three different groups following the OPT scoring procedure. The three groups were distributed according to the Common European Framework of Reference for Languages (CEFR) proficiency scale: Learners with a score in between 90 and 119 were categorized as "elementary" (A1-A2), scores between 120 and 149 were categorized as "intermediate" (B1-B2), and with a score higher than 150 as "advanced" (C1-C2), see Table 8 for descriptive statistics per proficiency group. The aptitude scores (all subtests of LLAMA and LLAMA global score) were recoded with a median split into lower and higher aptitude groups, see Table 9 for descriptive statistics per LLAMA subtest.

Table 8. Scores per proficiency group

	Proficiency score (Max: 200)			Pre-test score (Max:54)			Post-test score (Max: 54)		
	Mean (SD)	Min	Max	Mean (SD)	Min	Max	Mean (SD)	Min	Max
A1-A2 (n=14)	107.71 (8.48)	92	118	10.71 (5.69)	2	20	15.57 (6.02)	6	28
B1-B2 (n=32)	136.16 (9.42)	120	149	21.16 (8.41)	7	35	30.44 (9.01)	12	44
C1-C2 (n=23)	163.65 (10.21)	150	183	35.48 (5.63)	28	45	41.87 (5.45)	31	51

Table 9. Descriptive statistics per LLAMA subtest and aptitude group

		LLAMA B (Max: 100)				LLAMA D (Max: 100)				LLAMA E (Max: 100)				LLAMA F (Max: 100)			
		n	Mean (SD)	Min	Max	n	Mean (SD)	Min	Max	n	Mean (SD)	Min	Max	n	Mean (SD)	Min	Max
Captions	All	39	62.82 (17.04)	35	100	39	27.69 (15.68)	0	55	39	85.64 (13.91	50	100	39	56.92 (22.49)	0	90
	Lower	24	52.71 (11.88)	35	65	18	14.17 (9.73)	0	25	26	88.46 (11.55)	50	90	17	37.06 (17.23)	0	50
	Higher	15	79.00 (11.88)	70	100	21	39.29 (8.98)	30	55	13	100 (0.00)	100	100	22	72.27 (11.09)	60	90
No captions	All	30	62.86 (17.02)	30	100	30	19.82 (13.77)	0	40	30	75.36 (24.45)	0	100	30	47.50 (28.10)	0	100
	Lower	30	52.50 (10.60)	30	65	20	12.37 (9.62)	0	25	23	68.64 (24.74)	0	90	18	30.56 (23.12)	0	100
	Higher	75	81.50 (7.47)	75	100	10	35.56 (4.64)	30	40	7	100 (0.00)	100	100	12	74.17 (15.64)	60	100
Both groups	All	69	62.84 (16.90)	30	100	69	24.40 (15.31)	0	55	69	81.34 (20.06)	0	100	69	52.99 (25.22)	0	100
	Lower	43	52.62 (11.22)	30	65	38	13.24 (9.59)	0	25	49	73.96 (19.21)	0	90	35	33.71 (20.44)	0	50
	Higher	26	80 (9.01)	70	100	31	38.17 (8.03)	30	55	20	100 (0.00)	100	100	34	72.94 (12.68)	60	100

First, a GLM was fitted with the post-test scores as the dependent variable; group (captions vs. no captions), proficiency (elementary, intermediate or advanced), WM score, and aptitude group were included as fixed effects; and pre-test score was included as the covariate. Interactions were also introduced between group and proficiency, between group and WM scores, and between group and aptitude scores. The aptitude scores showed no significant main effect in this model (F (1,68) = .327, p = .570, ηp2 = .006), and neither did the interaction between group and aptitude (F (1,68) = .722, p = .399, ηp2 = .013), nor the interaction between group and proficiency (F (2, 68) = 2.212, p = .119, ηp2 = .072), nor the interaction between proficiency and aptitude (F (2,68)= 1.118, p=.333, ηp2=.035), nor the interaction between proficiency and WM (F (2, 68)= .388, p=.680, ηp2=.013). Non-significant fixed effects and interactions were removed from the model, and then the analysis was run again. The second model included the individual LLAMA subtests. The model did not return a significant main effect in the model for neither of the LLAMA subtests (p > 0.5). The final model (Model 1) included the pre-test scores as the dependent variable, group, OPT group, and WM score as fixed effects, pre-test score as the covariate, and an interaction between the group and WM scores. The GLM's script is provided in Appendix H.

Model 1 yielded a main effect for group (F (1,68) = 6.559, p = .013, ηp2 = .100) where the Captions group outperformed the no captions group, a main effect for proficiency (F (2,68)= 8.311, p = .001, ηp2 = .220), a significant main effect for WM (F (1,68) = 3.64, p = .023, ηp2 =. 084), and the pre-test scores (F (1, 68)= 169.001, p <.001, ηp2 = .741).

The Bonferroni pairwise comparisons between the proficiency groups showed that the intermediate proficiency group had significantly higher gains than the elementary

group (p = .004). Yet, no significant difference was found in the comparisons either between the elementary and the advanced groups or between the intermediate and the advanced groups (p > .05), see Table 10 for estimated marginal means per proficiency group.

Table 10. Model 1: Estimated Marginal Means (post-test) per proficiency groups

	All groups		Captions		No captions	
	Mean (SE)	95% CI	Mean (SE)	95% CI	Mean (SE)	95% CI
A1-A2	26.97 (1.61)	[23.74; 30.20]	29.85 (2.02)	[25.80; 33.89]	24.10 (2.01)	[20.06; 28.14]
B1-B2	32.47 (0.85)	[30.76; 34.17]	32.94 (1.13)	[30.67; 35.22]	31.99 (1.24)	[29.51; 34.47]
C1-C2	31.46 (1.42)	[28.60; 34.32]	32.15 (1.55)	[29.03; 35.27]	30.76 (1.90)	[26.94; 34.58]

The model also showed a significant interaction between group and WM scores (F (1,68) = 5.128, p = .027; ηp2 = .080). The WM scores were a significant predictor of the post-test score for the no captions group (β = .186, SE = .069, p = .008).

Finally, the assumption of homogeneity of regression slopes (Field, 2018) was met as the relationship between the pre-test scores (covariate) and post-test scores (dependent variable) were similar at different levels of group (independent variable, captions or no captions group) (F (1,68) = 1.497, p =.226, ηp2 = .025). The results of the GLM are shown in Table 11.

Table 11. Model 1: Results of the general linear model: Influence of fixed factors on constructions learning

	Df	Mean Square	F	*p*	ηp2
Group	1,68	128.400	6.559	.013	.100
Proficiency group	2,68	162.692	8.311	.001	.220
Working memory score	1,68	106.546	5.443	.023	.084
Pre-test score	1,68	3308.372	169.001	.000	.741
Group x working memory score	1,68	100.384	5.128	.027	.080
Model	6,68	1369.947	69.981	.000	.874
r^2 = 0.877					

As seen above, the effects of aptitude were found to be insignificant in Model 1 that included WM scores. However, we had predicted that aptitude may be a significant factor for lower-proficiency learners rather than for advanced learners (see Chapter 2). Therefore, another model was fitted without either the advanced group or working memory scores in case they were outweighing the effects of aptitude. Because of the small number of the elementary group (n = 14), the model was fitted with the elementary and intermediate level participants (n = 46).

The GLM (Model 2) included the post-test scores as a dependent variable, aptitude groups (lower or higher) per subtest (LLAMA B, D, E, F), group (captions, no captions), and the pre-test scores as a covariate. Model 2 indicated that the only subtest that had a moderate effect on the post-test scores was LLAMA F that measures grammatical inference ability (F (1,45) = 3.405, *p* = .072, ηp2 = .077), with a significant interaction between the LLAMA F group and the treatment group (F (1,45) = 5.913, *p* = .019, ηp2 = .126), see Table 12. The Bonferroni post-hoc analysis revealed that there was no significant difference between the treatment groups for the participants with higher LLAMA F (F (1, 45) = .065, *p* = .801, ηp2 = .002), but there was a significant effect of

group for the participants with lower LLAMA F (F (1, 45) = 10.820, *p* =.002, ηp2 = .209), with the captions group outperforming the no captions; the estimated marginal means are presented in Table 13. The model's script is provided in Appendix H. The assumption of homogeneity of regression slopes for Model 2 was met (F (1,45) = .185, *p* =.669, ηp2 = .005).

Finally, when the global, rather than componential aptitude scores were analysed, the results yielded no significant effect of aptitude (F (1, 45) = 1.355, *p* =.252, ηp2 = .034).

Table 12. Model 2: Results of the general linear model: Influence of fixed factors on constructions learning

	Df	Mean Square	F	*p*	ηp2
Group	1, 45	114.453	4.181	.047	.093
LLAMA F group	1, 45	93.226	3.405	.072	.077
Pre-test score	1, 45	3653.066	133.444	<.001	.765
Group x LLAMA F group	1, 45	161.875	5.913	.019	.126
Model	4, 45	1005.816	36.742	<.001	.782
r^2 = .761					

Table 13. Model 2: Estimated Marginal Means (post-test) per aptitude group

	All groups			Captions			No captions		
		Mean (SE)	95% CI		Mean (SE)	95% CI		Mean (SE)	95% CI
Lower	25	24.37 (1.05)	[22.23; 26.50]	17	27.86 (1.41)	[25.01; 30.72]	18	20.87 (1.57)	[17.69; 24.06]
Higher	21	27.26 (1.14)	[24.94; 29.58]	22	26.97 (1.57)	17.69; 24.06]	12	27.55 (1.66)	[24.19; 30.91]

5.4 Summary of the results

1) The participants had higher scores in the post-test regardless of group.
2) The captions group outperformed the no captions group.
3) Proficiency had a significant effect on the post-test scores: The intermediate group scored significantly higher than the elementary group. There was no significant advantage of the advanced group.
4) There was no significant interaction between the group and proficiency level; proficiency level did not affect the effectiveness of either of the viewing modes. However, there was a trend for the elementary group to perform better under the captions condition (Table 10).
5) Working memory capacity had a major effect when the TV series was viewed without captions.
6) Working memory was a more significant factor than aptitude in this study. When analysing the effects of aptitude without working memory scores and advanced level group, a moderately significant effect of LLAMA F was observed.

7) Language learning aptitude (as measured by LLAMA F) had a significant effect on learners in the elementary and intermediate proficiency groups who watched the episodes without captions.

5.5. Discussion

The aim of the study presented in this chapter was to explore the effect of extensive audio-visual input on the learning of L2 constructions. To our knowledge, this is the first study of a sustained intervention that has focused on the learning of grammatical constructions from TV series comparing two conditions: viewing with and without captions.

5.5.1 The effects of viewing TV series on grammatical constructions learning

The first research question addressed the overall effect of viewing ten episodes of the TV series on L2 grammatical constructions learning. All participants, regardless of their experimental group (with or without captions) significantly improved their knowledge of target constructions, which provides evidence for the learning potential of audio-visual input. Previous research on grammar learning from audio-visual input had dissimilar outcomes. The studies by D'Ydewalle and Van de Poel (1999) and Van Lommel et al. (2006) yielded no significant results after a short non-instructed intervention. It should be mentioned that the amount of input was very different in these studies: The participants in the present study were exposed to ten episodes (227 minutes) of the TV series while the earlier studies' participants were exposed to the audio-visual input only once (10 minutes in d'Ydewalle & Van de Poel, 1999; 25 and 16 minutes in Van Lommel et al., 2006). This brief exposure to audio-visual input may have been insufficient to lead to grammar uptake, in contrast to the vocabulary gains usually

observed in short audio-visual input studies (Montero Perez et al., 2013). This difference in results may imply that, in the case of grammatical constructions, a greater quantity of audio-visual input is required for rule abstraction and learning to take place than in the case of lexical units. This is in line with Kusyk and Sockett (2012) who suggested that students receiving a greater quantity of audio-visual input are likely to be more familiar with frequently occurring constructions in the TV series.

5.5.2 The effects of the on-screen text

The second research question aimed to explore the role of captions, proficiency, WM capacity, and language aptitude on learning L2 grammatical constructions through audio-visual input. Regarding the role of captions, we found a significant difference between the captions and no captions groups, where those participants who were exposed to captions significantly outperformed those who were not. This result is in line with previous research (Cintrón-Valentín et al., 2019; Cintrón-Valentín and García-Amaya, 2021; Lee and Révész, 2020), that found grammar learning gains after a brief captioned audio-visual exposure. This suggests that captioned audio-visual input supports the learning of constructions better than uncaptioned audio-visual input because it provides supporting written text and balances the processing loads in verbal and visual input channels (Vanderplanck, 2016: 147). Similarly, other studies exploring grammar learning through captioned audio-visual input (Lee & Révész, 2018; Plotnikova, 2017) found a significant improvement in grammar production after non-enhanced captioned exposure. Contrastingly, Van Lommel et al. (2006) did not find significant grammar learning (without pre-teaching) from a short movie presentation with L1 subtitles. It is important to note that the studies that found evidence of grammar learning from audio-visual input used L2 captions whereas those that did not find significant grammar gains implemented

L1 subtitles. Though the number of studies is still very limited and drawing conclusions may be premature, the existing evidence suggests that captions may be more favourable for grammar learning than subtitles.

5.5.3 The effects of proficiency level

Regarding proficiency, the analysis yielded a main effect of proficiency on constructions learning, and the superior gains of the intermediate over the elementary proficiency group. However, no significant difference was shown in the elementary and advanced groups' scores, nor the intermediate and advanced proficiency groups. In relation to the lower gains of the elementary proficiency group, the results are in line with Danan's (2004) idea that viewers need to reach a certain level of proficiency is take advantage of captioned videos exposure, as lower-level learners may have a limited processing capacity in the L2. Similar results of lower proficiency participants being outperformed by the higher proficiency peers have been observed in the studies on grammar (Plotnikova, 2017) and vocabulary (Pujadas & Muñoz, 2019; Suárez & Gesa, 2019). Conversely, there was no evidence of a proficiency advantage for the advanced proficiency group, whose gains failed to outperform either of the lower-level groups. A plausible explanation for this finding is that the advanced proficiency participants had less room for learning than the other groups as they were already familiar with many of the target constructions in the input. As a consequence of the informal nature of the TV series, where mostly high frequency grammatical constructions are present, advanced learners may have received a lower quantity of novel input than lower-level participants.

It is also worth mentioning that although the results showed no significant interaction between the group and proficiency level, an interesting trend can be observed

for the elementary group (see Table 10). There was a tendency for the elementary level participants in the captions group to perform better than the same proficiency level students in the no captions group. Even though the difference of 5 points did not reach significance probably due to the small number of participants in the elementary group (7 participants in each treatment group), this direction of captions supporting the elementary proficiency learners warrants further research. It would be crucial to establish whether certain levels of proficiency require particular viewing modes (i.e. captions, enhanced captions, keyword captions, no captions) for learning grammar from audio-visual input.

To summarise the proficiency results, even though all participants in our study, regardless of proficiency level, significantly improved their performance after the intervention, viewing episodes from a TV series seemed a more valuable source of input to learn target grammatical constructions for the intermediate proficiency students. It ties in well with the idea that captions become the most beneficial when viewers reach a certain proficiency level when the content of the audio-visual input is not too easy or difficult for the learner (Gass et al., 2019). However, more research is needed where input varies in difficulty (e.g. different genres in Suárez et al., 2021) and where participants also vary in proficiency level and experience with exposure to audio-visual material.

5.5.4 The effects of working memory capacity

As regards WM, the interaction between WM scores and group uncovered a greater reliance on WM skills by the no captions group than by the captions group.

WM capacity was more crucial to learn the target grammatical constructions for the participants who were exposed to uncaptioned audio-visual input. This ties in well with the assumption that captions have the potential of easing the cognitive load on the WM

as the information is distributed among the visual, auditory, and textual channels (e.g. Frumuselu et al., 2015; Vanderplank, 2016).

The results are the first to confirm a neutralising effect of captions on WM limitations (Wiley et al., 2014) by counter-balancing the differences in WM capacity (Gass et al., 2019). Montero Perez (2020) also uncovered that WM capacity was a significant predictor of learning pseudowords from uncaptioned audio-visual input. However, the researcher could not unearth the compensatory role of captions as her study did not have a captioned video group. This is also in line with Suárez et al. (2021) where no effect of working memory capacity was observed when the participants were exposed to captioned videos. The results of the present study suggest that the use of captions in that study may have neutralised any visible effects of WM. It may be suggested that WM is a significant factor for watching without captions because it is a more challenging task for a language learner than watching with captions. The no captions group may have needed to put in more cognitive effort to acquire grammatical constructions without the extra support provided by captions, while the captions group relied less on their WM as the availability of the on-screen text made the process easier for them.

The results of this study are in contrast with Lee and Révész (2021) where a positive effect of phonological short-term memory for captioned groups was observed. One of the explanations could lie in the differences between the working memory measurements used in the two studies. It is possible that the results may have yielded a positive effect of phonological short term memory for the captions group had we used an auditory test. Instead, the present study only used a reading span test. Secondly, extensive captioned viewing in this study (compared to the short clips in Lee & Révész, 2021) may have

neutralised the working memory capacity differences within the students in the captions group.

Finally, Gass et al. (2019) proposed that learners need to reach a certain proficiency level to activate the WM, as there was a tendency in their results for lower WM capacity participants to read captions more (as measured by eye-tracking). However, the results of the present study run counter to this assumption, as while proficiency level itself played a role, there was no interaction between WM capacity and proficiency in the current study. This suggests that the participants' WM activation did not depend on their proficiency level.

5.5.5 The effects of the language learning aptitude

Regarding aptitude, the results of the first model analysis showed that it did not have a significant effect on grammatical constructions learning. However, once we removed the advanced proficiency group and WM capacity scores from the model and analysed the effects of aptitude on the post-test scores, an interesting result was uncovered. It was found that LLAMA F scores had a moderately significant effect on the post-test scores, suggesting that those with higher analytical ability to infer and learn a grammatical structure performed better than those who had lower scores in LLAMA F. This finding goes in line with Winke (2013), who proposed that learners with higher proficiency levels rely less on their language learning aptitude. Indeed, the tendency for language learning aptitude to affect the post-test scores only appeared once the advanced learners' scores were removed from the analysis. The results are also in accordance with Feng (2022) where intermediate level learners with higher language learning aptitude outperformed those in the lower aptitude group. It can be suggested that advanced learners did not turn to their language aptitude while viewing the TV series, while elementary and intermediate

levels relied on it more. While this is a noteworthy finding for the area of audio-visual input, more studies with a comparison of lower proficiency learners should be designed, as the sample size of the elementary group in this study did not allow for a deep comparison between the proficiency and treatment groups.

As for the significant interaction between the LLAMA F group (lower and higher) and the treatment group (captions, no captions), the participants with higher language aptitude outperformed those with lower aptitude in the no captions group, while there was no difference between the aptitude groups in the captions group. This result indicates that the no captions group relied on their language learning aptitude significantly more than the captions group. This finding suggests that not only do captions neutralise variability in working memory, and it also extends the advantage of captions to those elementary and intermediate level learners who have lower grammatical inference ability.

The results in the present study only partially agree with the results from the studies on vocabulary learning through captioned audio-visual input where aptitude did not impact gains of word forms (Suárez & Gesa, 2019) or meaning recall (Muñoz et al, under review), but had a significant effect on gains of form-meaning pairing. Suárez and Gesa (2019) speculated that form-meaning pairing may have required greater cognitive involvement resulting in participants' language aptitude activation. They also suggested that the task drew more explicit focus to words' meanings than to their forms, and this difference in conditions could have affected the interaction with aptitude, as the LLAMA test largely measures explicit aptitudes (see Granena, 2013) and because aptitude is associated more with learning in explicit conditions (Li, 2015). Possibly, an explanation along those lines may account for the lack of reliance on aptitude within our participants in the captions group. In spite of the intervention taking place in the L2 classroom, the

lecturer never explicitly taught or focused attention on the target grammatical constructions. It is complicated to compare the results of the previous aptitude studies implementing audio-visual input, as the present study is the only[4] one which compared the captions and no captions groups. To confirm the findings in this study, further research should compare the effects of captioned and uncaptioned audio-visual input on learning various language features by participants with higher and lower language learning aptitude.

Finally, it is important to notice that the effects of aptitude were significant only when the working memory scores were not included in the model. It is possible that the abovementioned effects of working memory capacity may have overruled the role of aptitude in this study. Although these two factors were not correlated in the present sample and therefore measured different constructs, other researchers have found a stronger relationship between WM capacity and language learning aptitude (e.g. Wen, 2019). More research, especially with lower proficiency levels, should be conducted to tease apart the effects of aptitude and WM on learning from audio-visual input.

[4] Although the study by Feng (2022) had captions and no captions groups, it did not include an interaction between the aptitude groups and treatment groups.

Chapter 6

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

6.1 Introduction

The first study of this dissertation explored the effects of captioned and uncaptioned audio-visual input, and learner-related factors on learning 27 frequently occurring constructions in the target TV series. The study presented in this chapter extends the findings of Study 1 by addressing three constructions-related factors: constructions-type, frequency, and recency. Along with comparing the effects of captioned and uncaptioned conditions, the intervention also implements an additional captioning mode: Textually enhanced (TE) captions.

In the study presented in this chapter, the construction type is operationalised as three groups, based on Fried (2015), as described in Chapter 4 : fully-filled, partially-filled, and fully-schematic. Frequency is measured by number of times target constructions appear in the chosen TV series episodes. Finally, recency of occurrence is operationalised as test recency yielded by the comparison between learning gains on the immediate and delayed post-tests. Caption support is operationalised as three groups varying in the type of on-screen text: captions, textually enhanced captions (raised salience mode), and no captions.

This chapter is concerned with the following research questions (see Chapter 3 above):

Research Question 3: To what extent does L2 construction learning from audio-visual input depend on caption support, construction type, frequency, and recency?

Research Question 4: In what ways do these factors interact in L2 construction learning from audio-visual input?

6.2 Methodology

6.2.1 Participants

In order to control for proficiency differences, this study did not include the results of all 141 participants in this dissertation. Only the participants with the proficiency levels between B1 and C1 (intermediate to advanced, with a mean of B2) were included in the study, leaving 112 participants (see Table 14). Learner's proficiency level was shown to be a significant factor in the previous chapter. As the intact classes in this dissertation comprised of students with various proficiency levels (from A1 to C2), it was necessary to control for the language proficiency because the analysis (see below) was construction-based and participants' level could not be included as a factor.

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 post-tests; it was included to account for a possible testing effect from the immediate post-test (see Chapter 4). The viewing intervention for the captions and the no captions groups took place during the first

semester of the data collection (autumn 2018), and for TE captions and captions (-) during the second semester (autumn 2019).

6.2.2 Textual enhancement

The caption file was extracted from the TV series DVD. The application "Subtitle Edit" (Version 3.5.10; Olsson, 2019) was used to enhance target constructions for the TE captions group. The constructions were highlighted in yellow and bold (see Figure 3). Each episode featured 14 to 40 uses of target constructions, with 7 to 16 different target constructions per episode (see Table 4 in Chapter 4), representing about 4.5% of highlighted text in each episode. Only one construction was highlighted at a time; if there were two or more target constructions appearing on the screen at the same time, then the construction with lower frequency[5] was highlighted to avoid students splitting their attention between multiple enhanced constructions presented on the screen (Ayres & Sweller, 2014). The textually enhanced captions were burnt into the episodes with the "Handbrake" software (n.d.).

[5] The lower frequency constructions were selected for highlighting to facilitate their chances of being noticed.

Figure 3. A screenshot of the textually enhanced captions

6.2.3 Data analysis

This study analysed the pre-test, immediate post-test, and delayed post-test scores (see Chapter 4). The tests were scored dichotomously, 0 points were awarded if the answer did not elicit the correct usage of the target construction, and 1 point was awarded if all parts of the construction were used correctly. The students were not penalised for spelling mistakes as long as it did not hinder the meaning of the construction.

First, analysis of variance was run to see whether the treatment groups were comparable at the beginning of the intervention. Then, an independent samples t-test examined whether there was any difference between the captions group that performed the immediate post-tests and the captions (-) group that did not.

To answer the studies' research questions, 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 pre-test, immediate post-test, and delayed post-test 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 research question, fixed effects included captioning mode (captions, no captions, TE captions), construction type (fully-schematic, partially-filled, fully-filled), frequency of occurrence, and time (pre-test, immediate post-test, delayed post-test). Each construction was included as a random subjects effect in the model. 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). The script for the models is provided in Appendix H.

6.3 Results

6.3.1 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 14), or pre-test scores (F (3,108) = .311, p = .817) (see Table 15).

As mentioned above, the reason the Captions (-) group did not take the immediate post-test was to control for a possible testing or practice effect. To investigate this, an independent samples t-test was run with the delayed post-test 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 post-test itself was not observed. The captions (-) group was not included in the main analyses as these included the immediate post-test scores.

Table 14. 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 post-test

Table 15. 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

6.3.2 Research question 3: Factors explaining constructions learning

The third research 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 15 (the *n* for immediate post-test is smaller because captions (-) did not take this test).

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; the results are reported from Anova(model1) output. 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 16.

Table 16. 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	Pre-test	0.466 (0.038)	36.2	[0.387, 0.545]
	Immediate post-test	0.639 (0.038)	36.2	[0.560, 0.718]
	Delayed post-test	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 pre-test and immediate post-test (estimate = -0.172, SE = 0.012, p <.001) and between the pre-test and the delayed post-test (estimate = -0.174, SE = 0.012, p <.001), scoring on average 17% higher in both post-tests. There was no significant difference between the immediate and delayed post-tests' 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.

6.3.2 Research question 4: Interaction between construction learnability factors

The fourth research 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 results are reported from Anova(model2) output.

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 4). 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 4. 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 17. 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]

Regarding the pairwise comparison of the effects of captioning modes on immediate post-test 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 post-test 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 post-test scores (estimate = 0.000, SE = 0.021, p = 1.000). These results are summarized in Table 18.

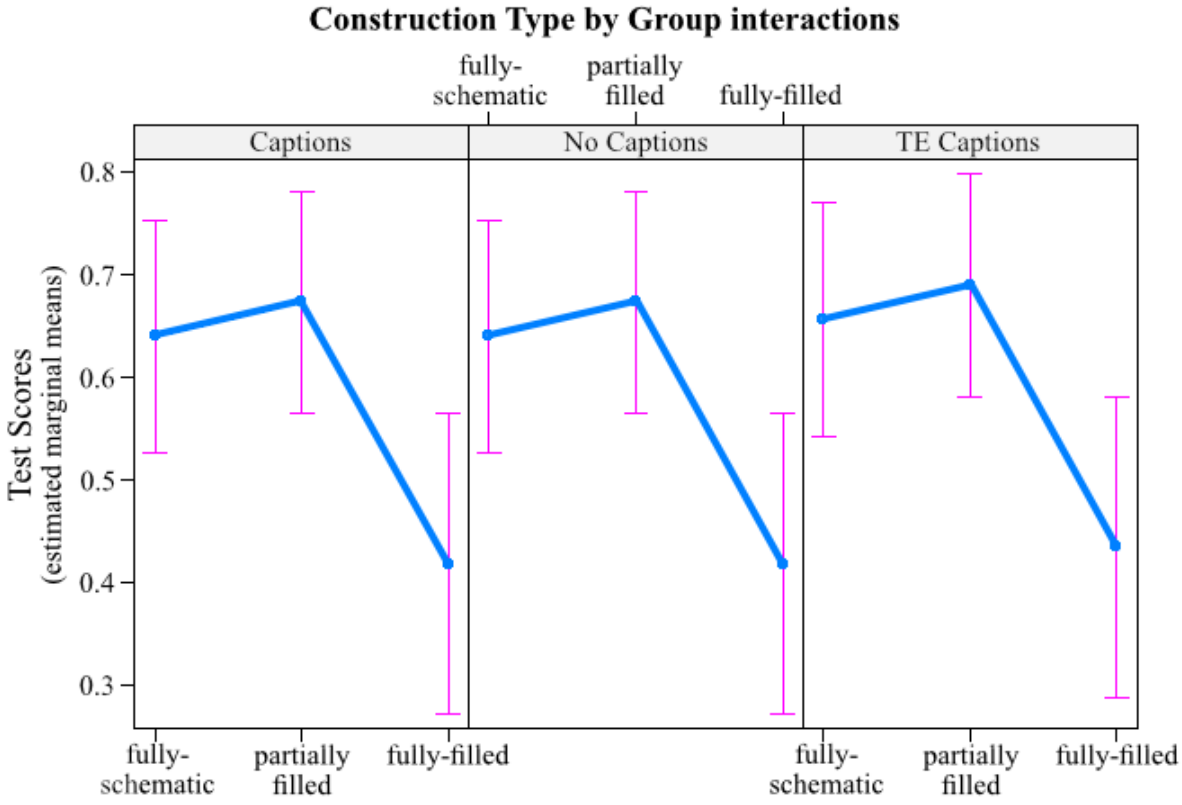

Figure 4. Test scores (divided by maximum possible score) by captioning mode group and construction type

Table 18. Coefficients of pairwise contrasts of group by time interaction, group comparison

Time	Group contrast	Estimate	SE	Df	t	*p*
Immediate post-test	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 post-test	Captions – no captions	0.055	0.021	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 19 and Figure 5) – 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 post-test (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 post-test (estimate = 0.058, SE = 0.021, p = .017). Finally, the no captions group did not have significantly different scores in the post-tests (estimate = 0.006, SE = 0.021, p = .945).

Table 19. Coefficients of pairwise comparison of group by time interaction, time comparison

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

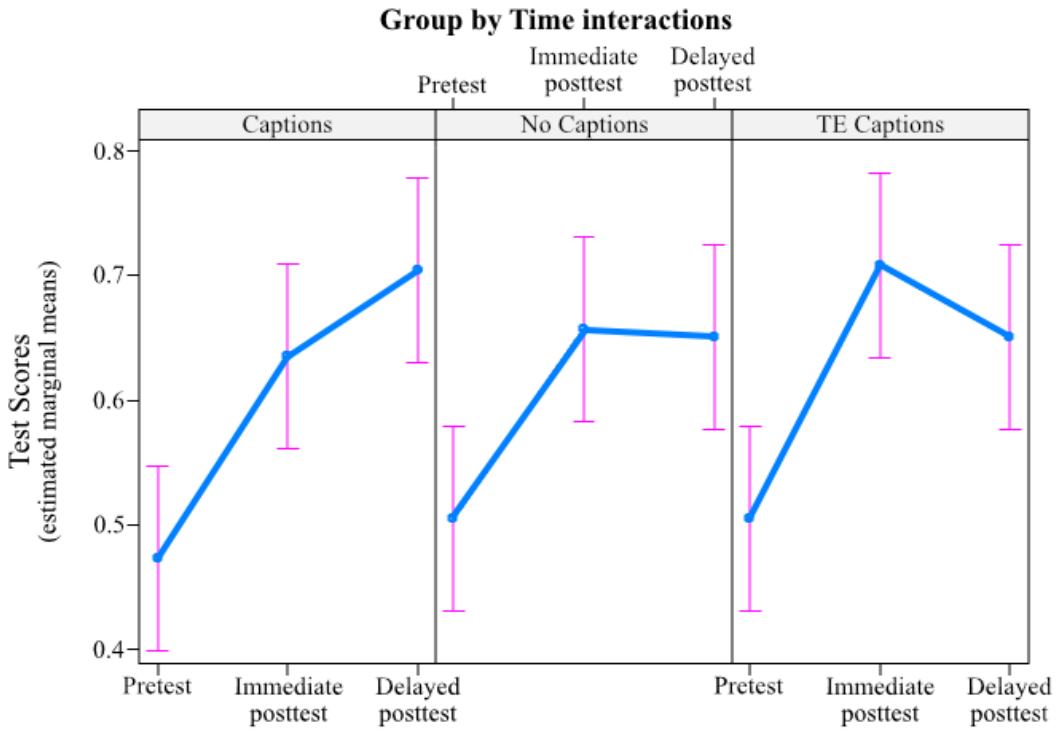

Figure 5. Test scores (divided by maximum possible score) by captioning mode and time of testing

6.4 Summary of the results

1) All the participants, regardless of the viewing group, had higher scores in the immediate and delayed post-tests than in the pre-test.

2) In the immediate post-test 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.

2) Contrariwise, in the delayed post-test, the captions group demonstrated the greatest scores, and the significant difference disappeared between the TE captions and no captions groups.

3) Regarding construction type, the partially-filled and fully-schematic constructions were learnt significantly better than fully-filled constructions.

4) Frequency of occurrence did not have a significant effect on learning of either of the groups.

5) As for test recency, the captions group had significantly higher scores at the moment of the delayed than of the immediate post-test; the TE captions group had higher scores in the immediate post-test; and the no captions group's scores did not differ significantly between the tests.

6.5 Discussion

The study presented in this chapter was designed to investigate the effects on grammatical constructions 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 target constructions. 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).

The third research question of this dissertation explored whether captioning mode, construction type, frequency of construction occurrence, and recency affected learning of the target constructions. The fourth research question focused on the ways in which these factors interact in L2 construction learning from audio-visual input.

6.5.1 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 post-tests. 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.

6.5.2 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. where elementary proficiency students were included in the analysis[6]. We may hypothesise that lower proficiency students exposed to uncaptioned 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.

[6] The elementary proficiency students of Study 1 (see Chapter 5) were included in that study.

6.5.3 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 post-test showed that the captions group outperformed both the TE captions and no captions groups. This supports the previously demonstrated benefit of captioned over uncaptioned 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 post-test 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 target constructions 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 post-tests 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 post-test. 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 target constructions 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.

6.5.4 Interaction between captioning mode and construction learnability factors

Finally, the fourth research 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 the target constructions. 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 target constructions.

As for the test recency with group interaction, we compared scores from the immediate and delayed post-tests 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 post-test than in the immediate post-test, suggesting that the long-term benefit from exposure to captions, i.e., the cumulative amount of encounters with the captioned target constructions, 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 post-test than in the delayed post-test, and it had higher scores than the captions group in the immediate post-test; 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 post-test, but did not have higher scores in the delayed post-test.

This is in line with Ellis's (2015: 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 target constructions in the input (the constructions appeared from 17 to 40 times in a single episode) likely 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 target construction 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. It is possible that the participants reduced their attention towards the enhanced constructions as they encountered more of them throughout the viewing sessions or even single episodes (as in the case of written texts in Indrarathne et al., 2018). 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.

Chapter 7

Perceptions of learning from audio-visual input and changes in L2 viewing preferences and strategies

7.1 Introduction

The first two studies of this doctoral dissertation explored learning of grammatical constructions and observed significant learning gains, as demonstrated by the post-tests. The study in this chapter extends the observed effects of audio-visual input, and explores the learning from the viewers' perspective.

While the results of audio-visual input studies attest a significant effect of both on-screen text and proficiency on learning gains, there is scarce research on whether these factors affect viewers' feeling of learning, a variable that can shape learner's perceptions of the learning process and affect the learning outcome (Ellis, 2008). In addition, studies have been documenting what mode of audio-visual input (i.e. with L1 subtitles, with L2 captions, without on-screen text) language learners are more likely to be exposed to (e.g Muñoz & Cadierno, 2021), but there is a lack of research exploring whether L2 viewers switch from one viewing mode to another and what factors affect those changes. Additionally, proficiency and intervention mode might influence learners' attitudes towards audio-visual input, and therefore its use. Moreover, there is very little descriptive research exploring viewing strategies as a factor, and in the

research that exists there is a lack of studies exploring this factor directly, rather than as a mere additional description, and especially through the lens of proficiency.

The study presented in this chapter extends the area of audio-visual input research by investigating the feeling of learning from different captioning modes by students with varying proficiency levels. It also contributes to the scarce research area of changes in viewing preferences and explores those changes by comparing them before and after a longitudinal intervention. Finally, it adds to the limited research available on viewing strategies. This chapter addresses the following research questions:

Research question 5: To what extent is students' feeling of learning from audio-visual input affected by intervention viewing mode and proficiency.

Research question 6: To what extent do students' viewing preferences change over time, and does this change depend on proficiency and/or intervention viewing mode?

Research question 7: What is the use of viewing strategies by different proficiency groups before and after the intervention?

7.2 Methodology

7.2.1 Participants

The study included responses from 136 participants, this number was smaller than the total number of the participants (N=141) because five students did not complete one of the questionnaires. The participants' proficiency was from A1 to C2, with a mean of B2 according to CEFR levels (Council of Europe, 2001). The four classes were randomly assigned to three different viewing conditions: Captions (n=71), no captions

(n=27), and textually enhanced (TE) captions (n=38). In this study the two captions groups (see Chapter 4) were merged because the grammar tests are not analysed in this chapter.

7.2.2 Materials

Participants' proficiency scores (OPT), responses to pre-course and post-viewing questionnaires, and class reflections were used in this study. The time period between taking the pre-course and post-viewing questionnaires was six weeks. The pre-course questionnaire on exposure to L2 audio-visual input contained questions on students' exposure to English media and viewing preferences (see Appendix A). The answers to four of these questions were analysed to address the study's research questions, see Figures 6, 7, and 8. The post-viewing questionnaire had the same format as the pre-course questionnaire, but also included two questions about participants' feeling of learning from *The Good Place* TV series. The participants were able to check all options that applied to them in both questionnaires.

Finally, after the intervention the participants were asked to write an essay with their reflections on the experience of watching the TV series in class (see Appendix F for the questions that students were asked to address). Many students elaborated on their experience with the TV series intervention during this task (see Appendix I), and therefore these reflections were included in the analysis to triangulate the quantitative findings.

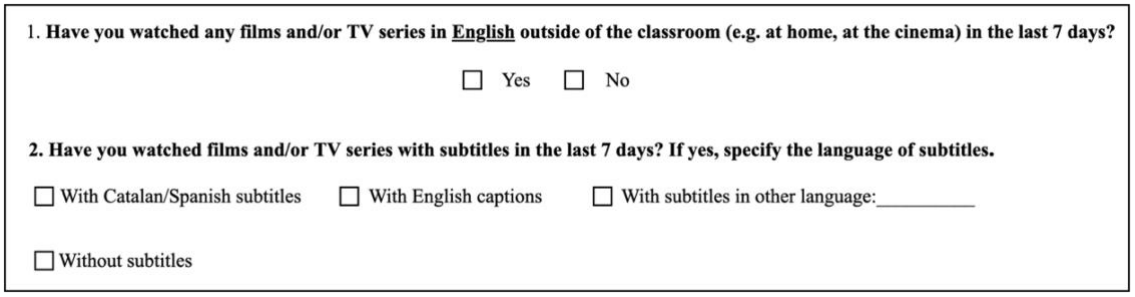

Figure 6. Questions on exposure to L2 original version audio-visual input, pre-course and post-viewing questionnaires

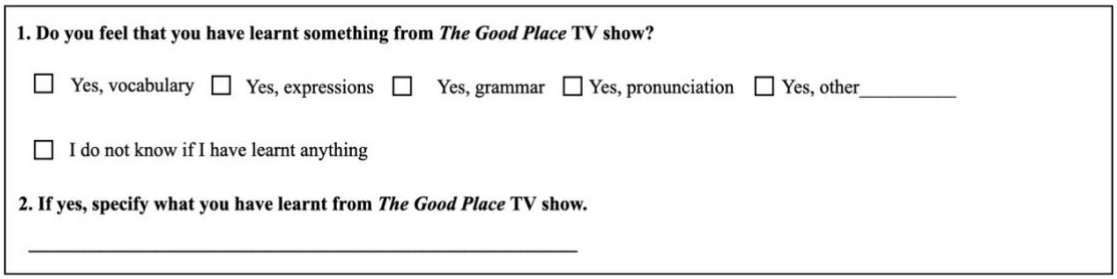

Figure 7. Question on feeling of learning from The Good Place in the post-viewing questionnaire

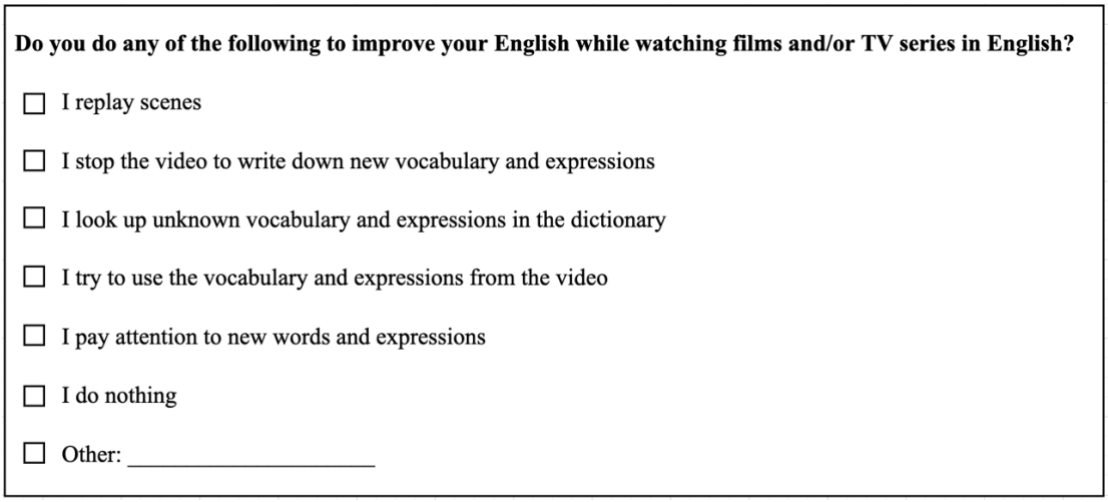

Figure 8. Question on viewing strategies while viewing outside of the classroom, pre-course and post-viewing questionnaire

7.3 Results

The data analysis was completed using SPSS software (Version 27). Proficiency level was operationalised as three groups based on the OPT scores, learners with a score between 90 and 119 were assigned to the A1-A2 elementary group, with the scores between 120 and 149 were assigned to the B1-B2 intermediate group, and with the scores higher than 150 were assigned to the C1-C2 advanced group. The descriptive statistics for proficiency scores are presented in Table 20.

Language feature was operationalised as vocabulary, expressions, grammar, pronunciation, and none (no feeling of learning). The students did not raise any language-related features in the "yes, other" option of the questionnaire, and therefore it was not included in the analysis. The use of captions was operationalised as "with L2 captions", "with L1 subtitles", and "without captions". Finally, learning strategies included "replay scenes", "stop the video to write down new vocabulary and expressions", "look up unknown vocabulary and expressions", "use the vocabulary and expressions from the video", "pay attention to new words and expressions", and "do nothing". The option "other" did not yield other strategies.

The preliminary analysis showed that there was no significant difference between the intervention groups' proficiency levels ($F(2, 133) = 2.242$, $p = .110$), nor their exposure to audio-visual input outside of the classroom before ($F(2,133) = .428$, $p = .653$) or after ($F(2, 133) = 2.218$, $p = .113$) the intervention. Therefore, the intervention groups in this study were comparable.

Table 20. Participants' proficiency scores (OPT)

	Mean (SD)	Min	Max
All participants	137.36 (19.82)	90	83
Captions	140.63 (19.31)	92	78
TE captions	132.52 (23.50)	94	83
No captions	132.55 (17.13)	90	68

7.3.1 Research question 5: Feeling of learning from the intervention

The first research question aimed to explore students' feeling of learning and whether it was affected by the intervention's viewing mode and students' proficiency.

To answer this research question a series of Binominal Linear Models for Repeated Measures were fitted in SPSS software. This analysis was chosen to address this research question because it allowed us to analyse all the language features in one model. The model included the aggregated binominal responses in one column for feeling of learning of the target features (yes, no) as a repeated measures dependent variable, and language feature (vocabulary, expressions, grammar, pronunciation, and none), intervention viewing group (with captions, without captions, with enhanced captions), and proficiency group (A1-A2, B1-B2, C1-C2) as independent variables. The model also included two interactions between intervention viewing mode and language feature, and proficiency group and language feature. The intervention viewing mode was not a significant factor ($F (2, 663) = .597$, $p = .551$), and neither was the interaction between the intervention viewing mode and language feature ($F (8, 655) = 1.424$, $p = .183$); therefore they were not included in the final model.

The final model yielded a significant effect of language feature (F (4, 665) = 21.235, $p < .001$) but not a significant effect of proficiency group (F (2, 665) = .979, $p = .376$). However, there was a significant interaction between proficiency group and language feature (F (8, 665)= 2.008, $p = .043$) on feeling of learning, suggesting that proficiency group itself could not explain students' feeling of learning.

Regarding the first significant factor, language feature, it can be seen in Figure 9 and Table 21 that expressions and vocabulary were perceived to be learnt the most, followed by no feeling of learning, pronunciation, and grammar. The post-hoc Bonferroni pairwise contrast revealed that there was no significant difference between the feeling of learning of expressions and vocabulary (t (665) = 1.804, $p = .258$). The analysis showed that vocabulary (V) and expressions (E) were considered to be learnt more than grammar (V: (t (665) = 6.501, $p < .001$), E: (t (665) = 8.758, $p < .001$)), pronunciation (V: (t (665) = 5.905, $p < .001$), E: (t (665) = 8.201, $p < .001$)), and none (V: (t (665) = 4.726, $p < .001$), E: (t (665) = 6.851, $p < .001$)). There was no significant difference in learning perception between pronunciation and grammar (t (665)= .907, $p = .589$), and there was no significant difference between no feeling of learning and pronunciation (t (665)= 1.049, $p = .589$), or grammar (t (665) = 1.852, $p = .258$).

Table 21. Estimated means of feeling of learning by language feature

	Mean (SE)	95% CI
Vocabulary	.473 (.047)	[.382; .567]
Expressions	.592 (.046)	[.501; .678]
Grammar	.095 (.034)	[.046; .185]
Pronunciation	.137 (.032)	[.086; .211]
None	.188 (.037)	[.125; .272]

To summarize this pairwise comparison, expressions and vocabulary were considered to be learnt the most, there was no difference between grammar and pronunciation feeling of learning, and the participants were more likely to report vocabulary or expressions learning than no learning at all.

Concerning the proficiency group and language feature interaction (see Figure 10), the comparison between proficiency groups revealed that grammar was perceived as learnt more by the elementary than the advanced group (t (665) = 2.590, p = .029). There was no significant difference between the rest of the group and language feature comparisons.

As for the interaction comparison within proficiency groups (see Table 22), the elementary group perceived more learning of vocabulary and expressions than pronunciation (V: (t (665) = 3.700, p = .002), E: (t (665) = 3.062, p = .018)), and no feeling of learning (V: (t (665) = 3.700, p = .002), E: (t (665) = 3.062, p = .018)). There was no significant difference between the vocabulary and expressions perceptions of

learning (t (665) = .531, p >.05), and no significant difference between grammar learning and other language features.

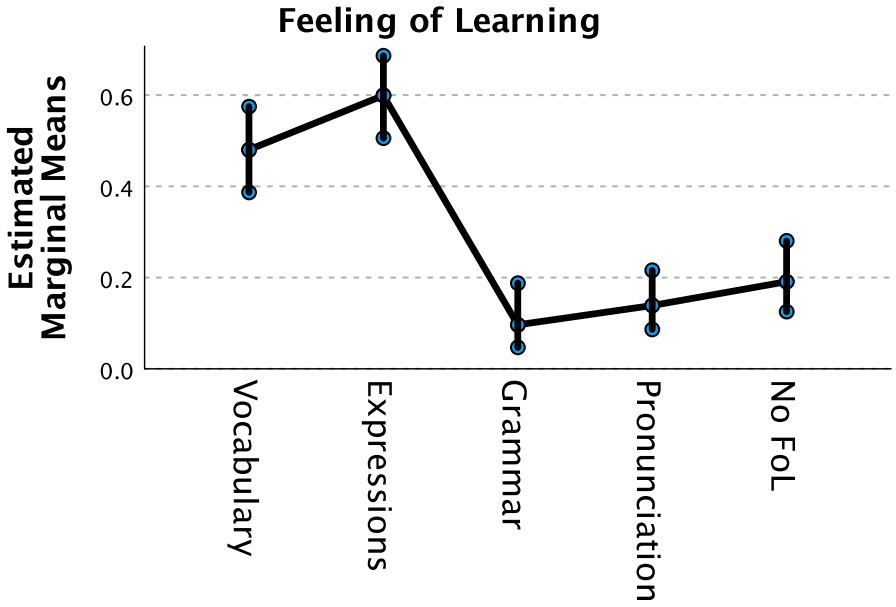

Figure 9. Estimates of Feeling of Learning for all participants

Table 22. Estimated means of feeling of learning by proficiency group and language feature

		Mean (SE)	95% CI
A1-A2	Vocabulary	.571 (.095)	[.384; .740]
	Expressions	.500 (.096)	[.321; .679]
	Grammar	.250 (.083)	[.123; .442]
	Pronunciation	.143 (.067)	[.054; .328]
	None	.143 (.067)	[.054; .328]
B1-B2	Vocabulary	.522 (.061)	[.403; .638]
	Expressions	.681 (.057)	[.561; .781]
	Grammar	.116 (.039)	[.059; .217]
	Pronunciation	.116 (.039)	[.059; .217]
	None	.159 (.045)	[.090; .267]
C1-C2	Vocabulary	.333 (.076)	[.203; .495]
	Expressions	.590 (.080)	[.430; .733]
	Grammar	.026 (.026)	[.004; .164]
	Pronunciation	.154 (.058)	[.070; .305]
	None	.282 (.073)	[.162; .443]

There was no significant difference between the feeling of vocabulary learning and expression for the intermediate proficiency group (t (665) = 1.917, p = .223). Vocabulary and expressions were perceived to be learnt more than grammar (V: (t (665) = 5.618, p <.001), E: (t (665) = 8.212, p <.001)), pronunciation (V: (t (665) = 5.618, p < .001), E: (t (665) = 5.618, p <.001)), and none (V: (t (665) = 4.806, p < .001), E: (t (665) = 7.232, p < .001)). There was no significant difference between the feeling of learning of grammar and of pronunciation (t (665) = -2.770, p > .05). Finally, the advanced group did not have a significant difference between vocabulary learning and expressions as the rest of the groups (t (665) = -2.324, p = .102). Both vocabulary (t

(665) =3.855, p = .001) and expressions (t (665) = 6.743, $p < .001$) were perceived to be learnt more than grammar. However, compared to the other proficiency groups, the advanced group's perception of vocabulary was not significantly stronger than pronunciation (t (665) = 1.867, $p = .187$), but they still thought that they learnt more expressions than pronunciation, like the rest of the groups (t (665) = 4.413, $p < .001$). Lastly, the advanced group tended to report no feeling of learning significantly more than grammar learning (t (665) = 3.320, $p = .007$).

The second question of the post-viewing questionnaire asked participants to give an example for each of the categories in which they had affirmed a feeling of learning. Many participants did not provide examples, or only provided examples for some categories. Those who responded mainly reported singular words and expressions (e.g. soulmate, to move on). Six students gave examples of grammar learning, five in terms of improving their "sentence order", and another reported that they had improved verb tenses. As for pronunciation, those who responded mostly wrote that they had learnt the difference between the American and British accents (one of the TV series characters had a British accent).

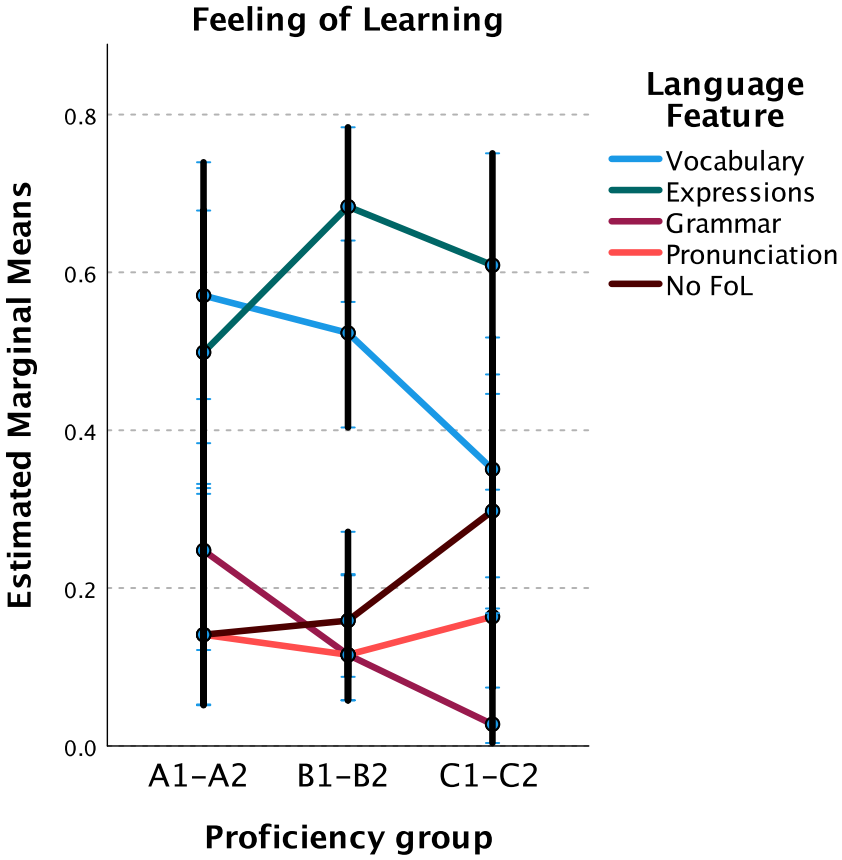

Figure 10. Estimates of Feeling of Learning per proficiency group

A qualitative analysis of students' reflections was also performed to triangulate the abovementioned results. The analysis suggested that many higher level proficiency students were concerned about the effectiveness of viewing TV series for language learning and suggested that some focus on form should take place as well if they want to improve their skills:

"I have to take notes and search for the difficult words to take advantage of the activity." (ID012, captions, C1-C2)

"I would say that watching content with subtitles didn't make magic and no one could learn English only from that. The learning process is not about watching something and hoping that will make you learn a foreign language, it is about trying hard to learn, to search and translate the words or expressions you don't know." (ID035, captions, B1-B2)

On the other hand, their peers from the elementary proficiency group were less critical and shared their positive beliefs about learning from audio-visual input:

"Thanks to watching the TV series now I have a wider range of vocabulary, and this is exactly the reason for which I have decided to watch more series in English from now on." (ID054, captions, A1-A2)

"I didn't trust that watching series in English would help me learn the language. But now I think it is one of the fastest and funniest ways to learn a language." (ID018, captions, A1-A2)

7.3.2 Learner's exposure to audio-visual input outside of the classroom

Research questions six and seven addressed students' exposure to audio-visual input outside of the classroom, and particularly participants' viewing preferences and viewing strategies outside of the classroom. The preliminary analysis included questionnaire data about participants engagement with the audio-visual materials before

and after the intervention, see Figures 11 and 12. At the beginning of the intervention 94% of students reported watching in the original version, and at the end of the intervention 90% of students reported watching L2 television as well. Figure 11 shows that the participants were engaged with L2 television weekly before and after the intervention.

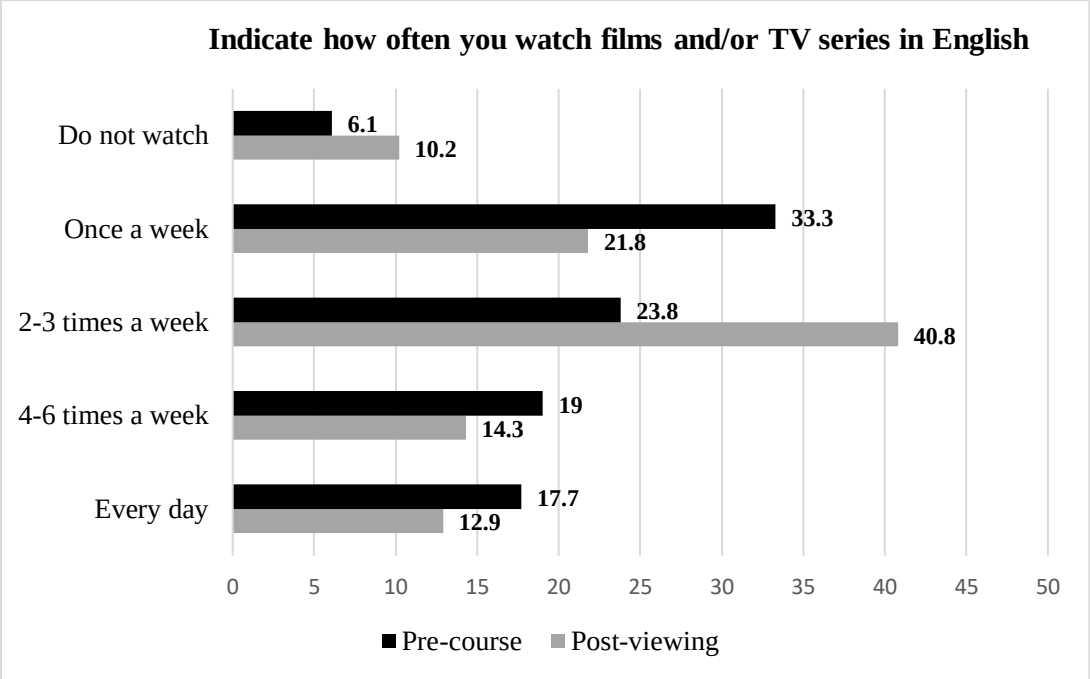

Figure 11. Amount of exposure to audio-visual input (in percentages)

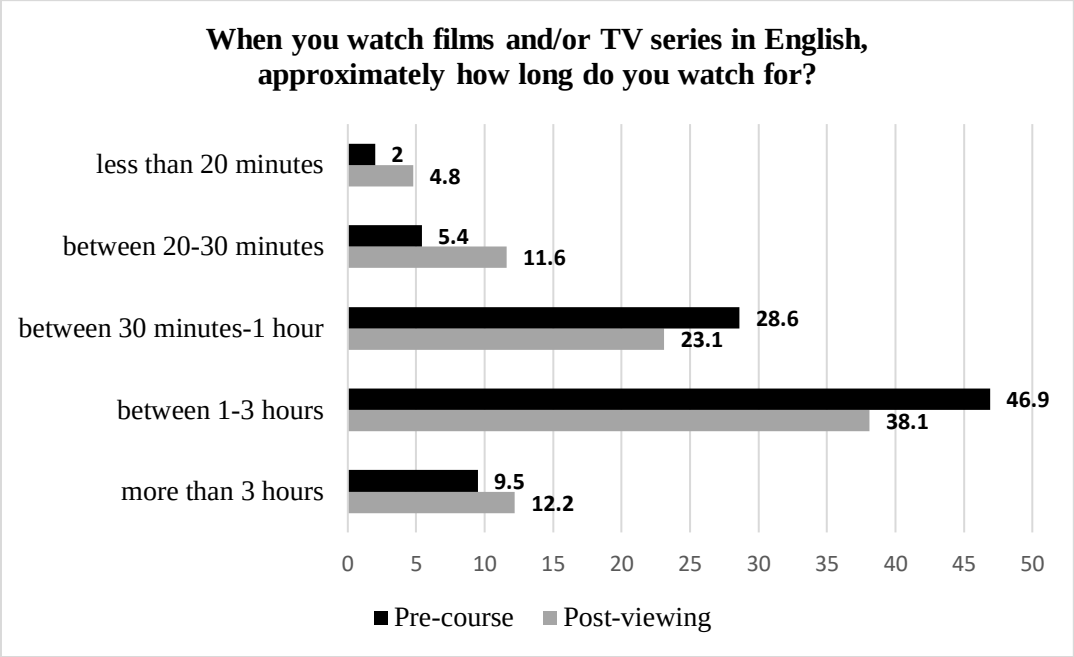

Figure 12. Length of single exposure to audio-visual input (in percentages)

Figure 12 indicates that the participants mostly spent between half an hour and three hours in a single viewing session.

When the participants were asked why they preferred to watch in the original version (Figure 13), the most popular responses were that they wanted to improve their English (63,9%) and that they liked to watch in the original version (76,9%). Several students explained in the "other" option that they particularly disliked the quality or the idea of dubbing.

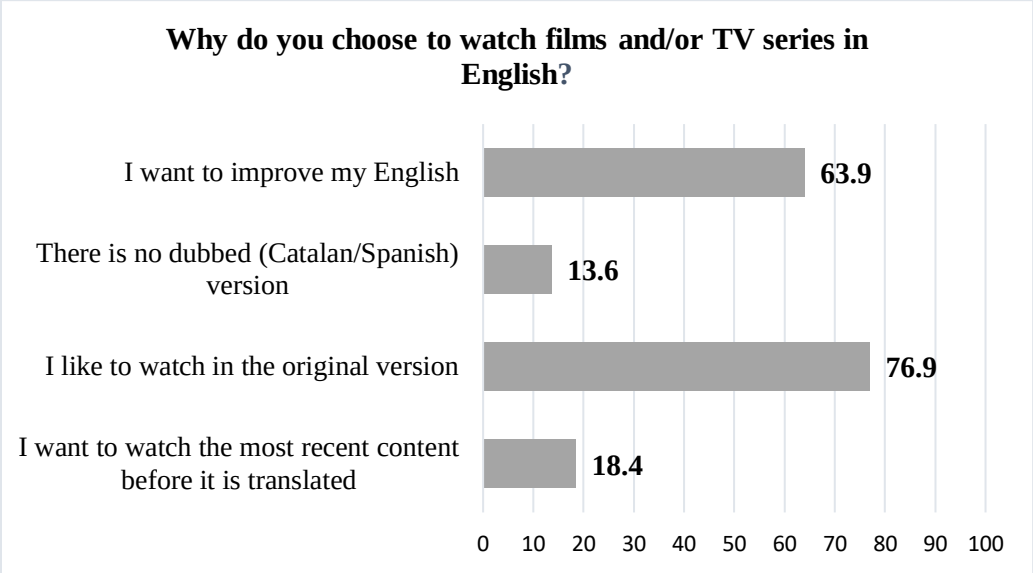

Figure 13. Reasons behind viewing original version audio-visual input

7.3.3 Research question 6: Viewing audio-visual materials outside of the classroom

The sixth research question analysed the issue of captioning mode preference outside of the classroom and whether this preference changed over the period of in-class viewing intervention. We were also interested in whether the changes were affected by participants' in-class viewing modes and their proficiency levels. To answer the second research question, a series of McNemar nominal paired samples analyses were run. This test was chosen because it allows analysis of the changes in dichotomous variables at two different points of time.

The answers to the survey question that asked participants if they watched media in English in the last week were analysed. If the participants chose a yes option, they then had to choose in what mode they watched the original version media (with L1

subtitles, with L2 captions, without subtitles or captions). The answers to the same question at the beginning and at the end of the intervention were compared.

First, the analysis was run with the answers from all the participants who reported watching English media in the last seven days regardless of the intervention viewing mode or proficiency group. The proportions of viewing preferences are presented in Table 23. The habit of viewing with L1 subtitles did not change (χ^2 (1) = 2.857, p = .091). The proportion of L2 captions users decreased at the end of the intervention (χ^2 (1) = 66.329, p < .001). Finally, the proportion of viewers without captions or subtitles significantly increased (χ^2 (1) = 21.729, p < .001).

In order to see whether these changes were affected by either the intervention viewing mode or participants' proficiency levels, the McNemar tests were run again. First, the effects of intervention viewing mode were considered. The Captions group's proportion of students who watched with L1 subtitles did not change (χ^2 (1) = .762, p = .383), while watching with L2 captions significantly decreased (χ^2 (1) = 35.220, p <. 001). The proportion of viewers watching without captions significantly increased (χ^2 (1) = 10.618, p = .001). As for the TE captions group, there was a tendency to watch less with L1 subtitles (χ^2 (1) = 3.200, p = .063), and there was a significant decrease of watching with L2 captions (χ^2 (1) = 19.048, p < .001); and a significant increase of watching without captions (χ^2 (1) = 11.250, p < .001). Finally, the no captions group did not change their preference regarding viewing with L1 subtitles (χ^2 (1) = .000, p > .05). Watching with L2 captions significantly decreased from (χ^2 (1) = 8.643, p = .002),

and the proportion of students viewing without captions did not change significantly (χ^2 (1) = .563, p = .454).

Table 23. The proportions of viewing preferences before and after the intervention

Group	Time	With L1 subtitles	With L2 captions	Without on-screen text
All participants	before	56%	78%	27%
	after	47%	11%*	61%*
Captions group	before	53%	77%	29%
	after	45%	11%*	64%*
TE captions group	before	59%	68%	15%
	after	45%	6%*	62%*
No captions group	before	51%	70%	38%
	after	54%	21%*	55%
Elementary group	before	81%	38%	24%
	after	67%	0%*	39%
Intermediate group	before	55%	71%	19%
	after	45%	9%*	68%*
Advanced group	before	44%	84%	41%
	after	38%	22%*	65%

**a statistically significant change before and after the intervention*

The last series of McNemar tests explored whether these changes in viewing preferences were affected by students' proficiency levels. For the elementary group, there was no significant difference in the proportion of L1 subtitles users (p = .375), but the intervention led to more participants watching significantly less with L2 captions (χ^2 (1) = 11.077, p < .001). There was an insignificant increase in viewing without captions

(p = .508). Similarly, the intermediate group did not change their L1 subtitles viewing habit (p = .210), and watched significantly less with L2 captions (χ^2 (1) = 32.237, p < .001). The proportion of viewers without captions significantly increased (χ^2 (1) = 21.441, p < .001). As for the advanced group, similar to the rest of the groups, there was no significant difference in viewing with L1 subtitles (p = .791), and there was a significant drop in viewing with L2 captions (χ^2 (1) = 19.360, p < .001). The proportion of viewing without captions did not change significantly (p = .124), but there was a tendency to watch more without captions.

These quantitative results are supported by students' end of the course reflections on their experience with the viewing intervention. Some lower-level students reported that they stopped viewing dubbed versions:

> *"Watching The Good Place helped me to change my habits of watching series in the language that has been filmed. However, I don't watch it with captions, but with subtitles in Spanish."* (ID125, enhanced captions, A1-A2)

While other elementary level students suggested that first they need to watch the episodes with L1 subtitles and then gradually switch to L2 captions:

> *"I think it is good to watch movies and series in the original language, although for me this is complex, since I need to make several views to understand, I have to see the chapters of the series first in the original language with Spanish subtitles and then with English subtitles."* (ID008, captions, A1-A2 level)

"After this experience, I think I have to watch more TV series and films in English with subtitles in Spanish and later with English captions to learn more new expressions." (ID016, captions, A1-A2 level)

Interestingly, although there was a significant drop in viewing with L2 captions, several students from captions group highlighted their intent to continue viewing with L2 captions:

"This experience helped me to see that I had to change my viewing style because I used to see series in English but with Spanish subtitles but now, after these exercises I decided to see the series with English captions and I think it will help me with improving my English." (ID105, enhanced captions, B1-B2)

"I almost never watch anything with captions because I thought that I would not understand. As I saw that I can understand most of The Good Place, I started watching TV shows with captions." (ID134, enhanced captions, A1-A2)

Finally, the results show that no captions group's participants opted for viewing without captions after the intervention:

"Watching The Good Place has changed my learning habits, it gave me confidence to not use subtitles anymore. My listening is getting used to and I feel very confident about it." (ID068, no captions, C1-C2)

"Now I find it easier to watch a TV series or a film without subtitles and captions. Maybe I easily forget the words or the grammar used by the

characters, but I understand what they say, and I can see the plot clearly. I guess I have a "better ear" now." (ID091, no captions, B1-B2 proficiency group)

"I attempted to watch movies without subtitles. Since we did it in class with The Good Place, I realized I could do it at home because I did not really need the Spanish subtitles anymore." (ID 093, no captions, C1-C2)

"I depended on the Spanish subtitles before and only after leaving them I understood my true level". (ID060, no captions, C1-C2)

7.3.4 Research question 7: Viewing strategies and language proficiency

The final, seventh research question addressed the learning strategies applied by the participants while viewing videos outside of the classroom. The aim of this research question was to compare the viewing strategies by proficiency level before and after the intervention. When the participants were asked to share their reasons behind viewing original version audio-visual materials (Figure 13 above), one of the most popular responses was to improve their English (63,9%). The participants were also asked whether they had any strategies to learn from audio-visual materials. As can be seen from Figure 14, the majority of the participants engage with audio-visual materials actively in order to improve their English. The most popular viewing strategies were paying attention to new linguistic content, trying to use the vocabulary and expressions from the video, and replaying scenes. The least popular strategies were stopping the

video to write down new vocabulary and expressions, and doing nothing. Although the participants mainly reported actively engaging with the language content of the videos, an increase in doing nothing while viewing L2 television may be observed in the post-viewing questionnaire responses. Additionally, an overall trend to engage less actively with audio-visual input could be observed.

To see whether this tendency was significant, the McNemar nominal paired samples analyses were run as the outcome variable was dichotomous ("yes" or "no") at two different points in time ("pre-course" and "post-viewing"). The results indicated that the proportion of students who implemented the strategies of replaying the scenes and stopping the video to write down new vocabulary and expressions did not change (p = .643; p = .424, respectively). The proportion changed significantly for the rest of the strategies with a decrease in looking up unknown vocabulary and expressions (χ^2 (1) = 8.654, p = .003), trying to use the vocabulary and expressions from the video (χ^2 (1) = 23.592, p < .001), paying attention to new words and expressions (χ^2 (1) = 8.205, p = .004), and with an increase in doing nothing while watching audio-visual materials (χ^2 (1) = 8.036, p = .005).

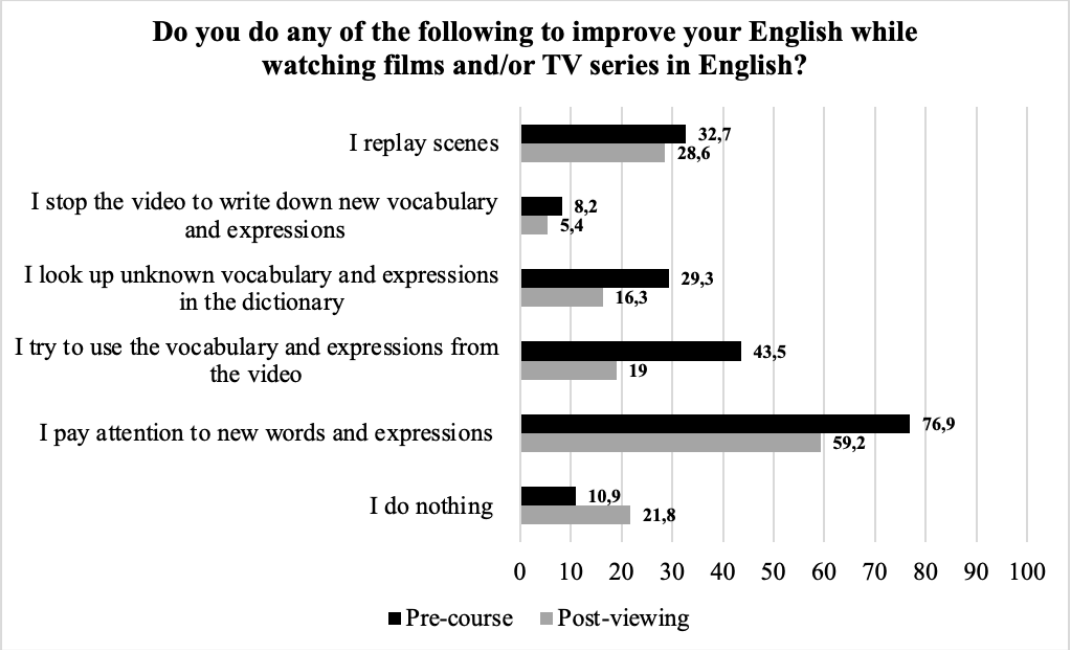

Figure 14. Viewing strategies of all participants before and after the intervention (in percentages)

In order to explore whether proficiency level had any effect on the viewing strategies, the questionnaire responses were analysed by proficiency group (elementary, intermediate, advanced), see Figure 15. A series of McNemar tests were run to observe whether the proportion of used strategies changed over time (Table 24).

The results indicated that the intermediate and advanced groups became more passive while viewing audio-visual materials as there was a significant change in the use of three strategies: Stopping the video to write down new vocabulary and expressions, looking up unknown vocabulary and expressions, and trying to use the vocabulary and expressions from the video in the intermediate group; and looking up unknown vocabulary and expression, trying to use the vocabulary and expressions from the video, and paying attention to new words and expressions in the advanced group.

Table 24. Changes in the viewing strategy use before and after the intervention

Strategy	Elementary group	Intermediate group	Advanced Group
Replay scenes	$p = .754$	$p = 1.000$	$p = .424$
Stop the video to write down new vocabulary and expressions	$p = .125$	$p = .031$	$p = .500$
Look up unknown vocabulary and expressions	$p = 1.000$	$p < .001$	$p = .002$
Try to use the vocabulary and expressions from the video	$p = .070$	$p < .001$	$p = .003$
Pay attention to new words and expressions	$p = .039$	$p = 1.000$	$p = .002$
Do nothing	$p = .125$	$p = .227$	$p = .109$

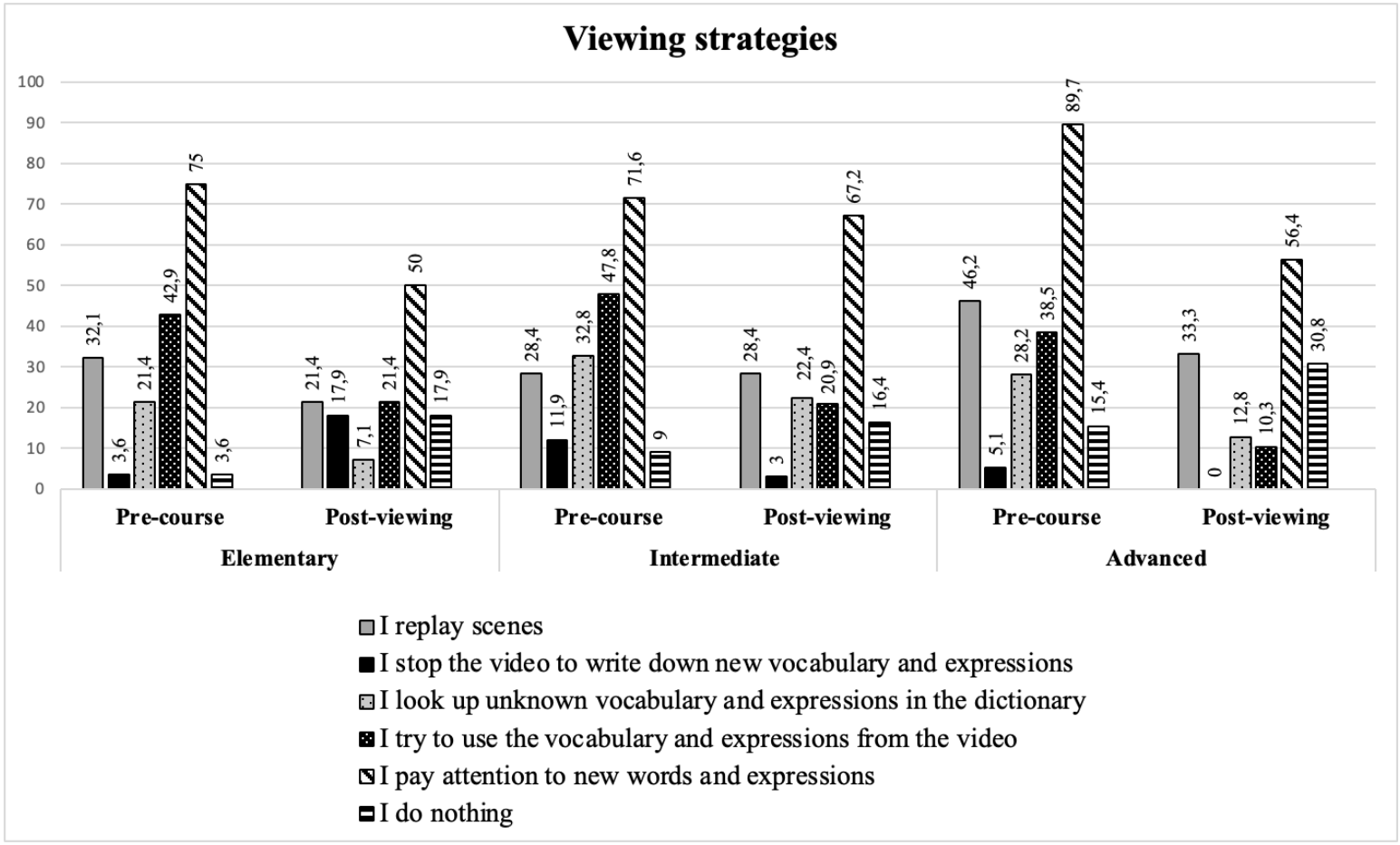

Figure 15. Percentages of strategies used before and after the intervention per proficiency group

The elementary group, on the other hand, significantly decreased their usage of only one strategy, paying attention to new words and expressions. The significant increase in doing nothing while viewing (Figure 14 above) was not attributed to proficiency level; however, it can be observed in Figure 15 (see above) that advanced proficiency viewers did nothing while viewing more than the lower proficiency students.

To explore whether there was an association between the proficiency level and strategies' use, a series of Chi-square tests were run (Table 25). The only significant association was found in the post-viewing data for stopping the video to write down new vocabulary and expressions (χ^2 (2) = 12.926, p =. 002), with the elementary group applying this strategy more than the rest of the participants.

Table 25. Association between the strategies use and proficiency level: Results of the Chi-Square test for association

Strategy	Pre-course	Post-viewing
Replay scenes	χ^2 (2) = 2.702, p = .254	χ^2 (2) = .613, p = .736
Stop the video to write down new vocabulary and expressions	χ^2 (2) = 2.694, p = .260	χ^2 (2) = 12.926, p = .002
Look up unknown vocabulary and expressions	χ^2 (2) = 1.354, p = .508	χ^2 (2) = 4.290, p = .117
Try to use the vocabulary and expressions from the video	χ^2 (2) = 1.474, p = .478	χ^2 (2) = 3.036, p = .219
Pay attention to new words and expressions	χ^2 (2) = 2.909, p = .233	χ^2 (2) = 4.530, p = .104
Do nothing	χ^2 (2) = 2.331, p = .312	χ^2 (2) = 2.264, p = .322

7.4 Summary of the results

1) There was no effect of the intervention viewing mode (captions, TE captions, no captions) on feeling of learning.

2) There was no difference between feeling of learning of vocabulary and expressions for all three proficiency groups, suggesting that vocabulary and expressions were considered to be learnt the most regardless of proficiency level.

3) The elementary group felt that they were learning grammar more than the advanced group. Also, there was a tendency for higher proficiency participants to have weaker feeling of grammar learning (see Figure 10). The advanced group also had a tendency towards reporting no feeling of learning at all.

5) The qualitative results of students' reflections suggested that elementary group participants were more likely to be positive about learning from the intervention, while higher proficiency students appeared to be more dubious.

6) The participants' habit of watching with L1 subtitles diminished (though the differences were not statistically significant) over the time period of this study, and most of the participants showed a pattern to shift from viewing with L2 captions to without captions.

7) There was no variation between the viewing modes groups, but there was a significant decrease in viewing with L2 captions for all three proficiency groups, and a significant increase in viewing without captions only for the intermediate group.

8) The qualitative analysis of students' reflections showed that elementary level participants realised that they are capable of watching in the original version, but with the L1 support first. The results also tentatively suggested that higher proficiency learners who watched without captions switched to this viewing mode daily.

9) Regarding viewing strategies, paying attention to the unknown vocabulary and expressions, and trying to use new vocabulary and expressions were the most reported strategies among all the participants, regardless of their proficiency level. The least popular strategies were stopping the video to write down new linguistic content, and doing nothing. The participants significantly decreased their use of strategies after the intervention. The elementary proficiency group reported stopping the video to write down new vocabulary and expressions significantly more than the intermediate and advanced groups.

7.4 Discussion

The study in this chapter was set out to examine how the participants perceived learning from an audio-visual intervention. It also analysed students' shift in viewing preferences before and after watching ten episodes of a TV series. The results were analysed in light of such factors as participants' intervention viewing condition (with captions, with textually enhanced captions, without captions), and proficiency levels. In addition, an analysis of viewing strategies and their association with students' proficiency level was performed. This section will discuss the significance of the obtained results for research questions five, six, and seven of this dissertation.

7.4.1 Feeling of learning from the intervention and its relationship with proficiency level

The fifth research question explored whether the feeling of learning from the audio-visual input depended on the intervention viewing mode and participants' proficiency. The results indicated that the intervention viewing mode did not affect feeling of learning. This result is surprising as previous research established that

captioned audio-visual input is objectively more beneficial than uncaptioned for various language features (see Montero Perez, 2022). However, several studies have shown that actual learning gains and feeling of learning do not always match (e.g. Sydorenko, 2010; Deslauriers et al., 2019). This lack of intervention viewing mode effect confirms that students may not always be aware of their learning progress or the fruitfulness of L2 captions use for their language learning. In her recent state of the art review of audio-visual input for L2 learning, Montero Perez (2022) suggested that learners' awareness of the benefits of audio-visual input and various on-screen text modes should be raised so that learners could fully engage with the input and benefit from it to a greater extent. Language instructors should take this suggestion and the present study results into account while informing their students about the benefits of various on-screen text modes and audio-visual input.

Regarding the language features, vocabulary and expressions were perceived as learnt the most, while pronunciation and grammar the least. This is in line with previous research mainly reporting perceived learning of vocabulary and expressions (e.g. Dizon, 2021). Grammar was considered one of the least learnt features, however, the results of Studies 1 and 2 with the same participant pool suggested that students in all the intervention groups had significant grammar gains, with the captions group outperforming the rest of the participants (in the delayed post-test, after viewing all ten episodes). One of the explanations of a low feeling of grammar learning in the present study may lie in the unconscious and incidental nature of uptake from audio-visual input (Vanderplank, 2016). It is possible that the participants were not aware of learning taking place, especially for such complex language feature as syntax. Primary school viewers in Avello (in preparation) received instructions on what counts as grammar before they

completed the post-viewing questionnaire, but grammar was still perceived to be learnt the least. Perhaps, with their greater maturity however, the adult participants in the present study would have recognised grammar learning more if they were instructed on what counted as grammar.

Another possible explanation of this contrariety in the results could also lie in the lack of feedback (Sydorenko, 2010), and no classroom opportunity to produce comprehensible output (Swain, 1985). It might be the case that exposure to audio-visual input is enough to obtain grammar gains, but a lack of output or explicit instruction could result in a weaker feeling of learning. Finally, the participants' metalinguistic awareness may have affected their responses, as they may, for example, have had a narrow view of what counts as "grammar". The responses to the open-ended question indicated that the participants' understanding of which grammatical structures they encountered was essentially limited to "sentence order", and may not have included such syntactic structures as reported speech, passive constructions, and conditionals that appeared in the TV series. If they fail to recognise parts of language they encounter as grammar, then it follows that they will not report having learnt grammar.

As for the proficiency groups, some interesting results were obtained for the elementary (A1-A2) and the advanced (C1-C2) groups. The lower proficiency participants reported a stronger feeling of grammar learning than their more advanced peers. Additionally, the qualitative analysis of students' reflections revealed that the elementary group was more positive about their overall learning outcomes than the participants with higher proficiency levels. In return, the advanced group was more likely to report no feeling of learning than the rest of the groups and was more sceptical about their progress and the educational value of the TV series. These results could be

explained by the rate of learning at different proficiency levels (Knight, 2018). The elementary level students had more room for improvement, more "low hanging fruit" in terms of relatively simpler words, phrases, and structures available, and consequently a higher feeling of progress. Once a learner reaches the "intermediate plateau", their progress decelerates, resulting in a weaker sense of progression. It seems that especially higher proficiency learners would have benefitted from explicit feedback and practice to see their progress, otherwise viewing TV series might be seen solely as a leisure non-educational activity (Vanderplank, 2019). The danger of students not perceiving audio-visual materials as a source of language learning lies in the psychology findings on investment of mental effort (Salomon, 1984). When learners consider an activity easy or entertaining (i.e., videos), they make less effort to learn from it, which consequently results in lower learning achievement. Also, a different type of materials with more infrequent vocabulary, expressions, and grammatical constructions, could have given the advanced learners a higher feeling of learning since they would have had more room for improvement.

7.4.2 Effects of intervention and proficiency on changes in viewing preferences

The sixth research question gauged a shift in preferred viewing mode (i.e., L1 subtitles, L2 captions, no on-screen text). We were particularly interested in how the intervention viewing mode (Captions, TE captions, and no captions) and participants' proficiency level (elementary, intermediate, advanced) could have affected those changes.

The results indicated that neither intervention group nor proficiency level caused a substantial change in viewing with L1 subtitles. It is possible that this change did not

reach significant difference as the participants might not have been able to completely switch from L1 subtitles. The reasons for this might lie in their lower proficiency, as reported by several elementary level participants in their course reflections, or external factors such as family and friends who could not follow the content without L1 on-screen support (Pujadas, 2019).

The biggest change in the viewing preferences was observed with the L2 captions mode. At the beginning of the intervention 78% of participants reported viewing with L2 captions, while by the end of the intervention only 11% of students claimed they viewed with L2 captions. The reasons for this significant switch for all intervention groups and proficiency levels may be various. The elementary group might have found that the intervention experience of watching with L2 captions regularly was too challenging. It was suggested by several elementary proficiency participants that while watching videos at home, they had to watch the videos several times to grasp the content, and they still relied on L1 subtitles. As for the more proficient participants, a natural shift from viewing with L2 captions to without captions was observed; the participants might have realised that they did not have to continue relying on captions anymore as they became more confident viewers throughout the intervention.

As for the effects of the intervention's viewing mode, the reason why all groups opted for viewing less with on-screen support could also lie in their raised confidence in viewing original version audio-visual input (Vanderplank, 2019). After a regular exposure to L2 captions, the captioning groups may have realised that they did not need the on-screen support anymore. Similarly, the no captions group had to adjust to viewing without any textual support, resulting in the creation of a habit of viewing more without captions, as reported by several participants. The results of this study do not support

previous studies' concerns about over-reliance on captions (e.g. Winke et al., 2010). The longitudinal nature of the present study allowed us to observe how students, starting from the intermediate proficiency level, pass the captioning phase and move on to uncaptioned input.

Finally, it is worth noting that although the analysis did not yield any increase in viewing with L2 captions, or switch from L1 subtitles to L2 captions, several individual responses provide evidence that some students benefited from the captioned intervention and took up a habit of viewing with L2 captions, suggesting that this viewing intervention inspired a few students to move from L1 subtitles to L2 captions.

7.4.3 Viewing strategies and proficiency

The final, seventh, research question of this dissertation addressed the viewing strategies applied by the participants while they watched original version audio-visual material outside of the classroom, and whether those strategies changed over time or/and were associated with the viewers' proficiency level.

The participants indicated that one of the main reasons they watched TV series and movies in English was to improve their language skills. It was confirmed by the results of the viewing strategies analysis as only 11% of students reported doing nothing while viewing, suggesting that the rest of the participants were actively engaged with the video materials to some extent. This active engagement mainly included paying attention to the unknown vocabulary and expressions, and trying to produce newly encountered vocabulary and expressions. The strategies that were employed the least were stopping the video to write down new words and expressions, and doing nothing. This is in line with the previous research (Alm, 2021), which has proposed that L2

viewers can be reluctant to interrupt the videos to write down new information as they would rather watch for enjoyment, even if they cannot understand every word.

Regarding the effects of language proficiency, our results do not fully go in line with other language learning strategies research (Oxford, 1989; Tang & Tian, 2015) suggesting that proficiency affects their use, and particularly that higher proficiency level students employ more strategies. However, this variance could be attributed to the difference in the range of strategies explored by Oxford (1989) from the ones applied in the present study. Our results indicated that proficiency was a differentiating factor for the use of one of the strategies, but that difference was not associated with the higher proficiency. Rather, it was the elementary group who reported stopping the video to write down new vocabulary and expressions more than the rest of the proficiency groups. Interestingly, this strategy is generally one of the least popular ones. Previous studies (e.g. Alm, 2021) showed that students did not want to pause the videos as they wanted to enjoy the viewing without interruptions and would rather understand the meaning of unknown words through the context. However, those students already had an intermediate level, and therefore could follow the videos and understand the most frequent vocabulary, as they probably had reached the vocabulary size of 3000 words necessary to understand original version audio-visual input (Webb & Rodgers, 2009). In the present study elementary students were more likely to use the pausing strategy. It is possible that the experience of watching in the original version was already challenging enough for the lower proficiency students, to the point that they could not properly follow and enjoy the content by merely watching. Therefore, they turned to this strategy more than the rest of the participants in order to comprehend the material by clarifying some key vocabulary. Higher proficiency students, on the other hand,

probably were able to decipher the meaning of the unknown words from the context, and did not have to apply the strategy of pausing the videos.

Finally, the finding that the participants implemented fewer viewing strategies at the end of the intervention could be connected to the fact that they watched less with L2 captions (see above). In case of the intermediate and advanced groups, the participants switched to viewing without captions. Watching without captions implies less engagement with the linguistic content of the videos, as it may be associated with low effort (Salomon, 1984; Vanderplank, 2019). As the higher-level participants switched to no captions viewing mode, it is also likely to indicate that they had become more confident in viewing for general comprehension and leisure. As they realized that they were capable of watching the videos without captions, they did not need to apply other engagement methods such as strategies that they were probably using before to scaffold themselves. Moreover, the viewing strategies might have become more difficult to apply, as the participants, for instance, did not see new vocabulary and expressions on the screen as often and it was easier to miss new language tokens with no textual support. In addition, a possible reason the participants opted for leisure viewing rather than active learning from audio-visual input is the fact that the post-viewing questionnaire was taken in the middle of the busy academic semester, a time when students' energy may have been consumed by classwork rather than self-study.

This finding, together with the discovery that students were switching to uncaptioned videos after the intervention, suggests that intermediate and advanced level participants felt more comfortable with the input and became viewers rather than learners. As for the elementary group, they also switched from viewing with L2 captions, however, for a different reason. The lower proficiency students had to rely on L1

subtitles as they realized they could not deal with the complex task of watching without L1 support. As mentioned above, the absence of L2 text on the screen does not facilitate noticing new vocabulary and expressions, and therefore minimized the use of active viewing strategies.

Chapter 8

Common Discussion and Conclusion

This doctoral thesis set out to explore the potential of original version audio-visual input on grammatical constructions learning. It aimed to answer seven research questions that were explored in separate studies in Chapter 5, Chapter 6, and Chapter 7. Study 1 addressed the role of captioned and uncaptioned input on grammatical constructions learning, and explored how this learning might be affected by such learner-related factors as proficiency, language learning aptitude, and working memory capacity. Study 2 extended the results found in the first study and explored the captioning mode differences and linguistic input-related factors that may affect learning from audio-visual input. In particular, Study 2 investigated the effects of different types of captioning (i.e. unenhanced captions, textually enhanced captions, and no captions), and examined whether such construction learnability factors as construction type, frequency of occurrence, and recency affect learning. Finally, Study 3 focused on the viewer's perspective of learning from audio-visual input and examined whether the feeling of learning and changes in viewing preference and strategies might be affected by proficiency level and intervention viewing mode.

This chapter will present an overview of the findings obtained in the three studies, discuss the studies' limitations, and provide future research directions, pedagogical implications of the results, and conclusions.

8.1 The main findings

The findings of the three studies of this dissertation support the potential of audio-visual input for grammar learning. The interventions led to significant learning of the grammatical constructions regardless of the viewing mode (captions, TE captions, no captions).

The results also provide evidence for the benefits of on-screen text for grammar learning. The groups who viewed the material with captions outperformed the no captions group in the post-test. The results also shed light on the comparison of two captioning techniques - captions, and textually enhanced captions - for prolonged viewing of TV series. The results support previous findings on the effectiveness of TE captions; however, the longitudinal nature of the experiment allowed us to observe that there was only a short-term benefit of TE captions. It seems that unenhanced captions lead to better long-term retention of the grammatical constructions. However, the viewing mode did not affect their feeling of learning from the participants' point of view. That is, captions groups (both unenhanced and enhanced) did not perceive the intervention as more helpful for their language learning than the no captions group.

The mediating role of proficiency on learning from audio-visual input was confirmed, with the intermediate proficiency group obtaining the highest gains. The findings also suggest that learners with an advanced proficiency level need more complicated content - containing lower frequency constructions - to have sufficient room for growth in their constructions knowledge. It also seems that advanced learners need extra feedback from their instructors or additional focus on form activities in order to perceive learning from audio-visual input positively. On the other hand, although the elementary group was the one who learnt the least from the intervention, they in fact

reported more feeling of learning, especially for grammar, than the advanced group. Regarding proficiency, it seems the sweet spot is when you are proficient enough to be able to follow the material, but not so proficient that the material contains little new language. Finally, although there was no significant interaction between the intervention group and proficiency group, the elementary students tended to learn more if they were in the captions group. This finding provides some evidence that intermediate proficiency is a threshold to be able to watch, comprehend, and learn without captions.

As for individual cognitive differences, the results bring to light the neutralising effect of captions on limitations in both working memory and language learning aptitude. It was found that high working memory was a marker for success when watching without captions, but it made no difference when watching with captions. Working memory capacity was a more important factor in this study than the language learning aptitude, however, once WM capacity and the advanced group's scores were removed from the model, a significant interaction between the grammar inference ability (as measured by LLAMA F) and group appeared. This interaction provided additional evidence of captions leveling the playing field, as the participants who watched without captions needed to rely on their aptitude more.

Regarding the constructions learnability factors, the partially-filled and fully-schematic constructions were learnt significantly better than the fully-filled constructions. This suggests that more restricted, fixed constructions are more difficult to uptake from audio-visual input (whether captioned or uncaptioned). Frequency of occurrence did not affect learning. This surprising finding may reflect the fact that frequency of occurrence might have a stronger effect on elementary proficiency learners. The elementary learners were not included in the constructions learnability

study of this thesis (Study 2), but were included and analysed in a related article using the same participants and materials as in Study 1, where it was found that there was an effect of frequency of occurrence only for the lower proficiency learners (Muñoz et al., 2021). In terms of test recency, the viewing groups went in different directions and demonstrated stronger short-term or long-term effects depending on the captioning mode (see above).

We move now to feeling of learning. Vocabulary and expressions were perceived to be learnt the most by all participants. The significant gains in the grammar post-tests were not represented in the participants' feeling of learning, suggesting that actual and perceived grammar learning do not coincide because grammar was perceived as learnt the least. This supports previous findings on learners not necessarily being aware of the learning taking place, and of the necessity to direct learners' attention towards the potential of audio-visual input for grammatical constructions learning. However, it is important to keep in mind that students' may have felt that they learnt so much more vocabulary relative to grammar, that they felt their grammar learning was inconsequential, and therefore not reported it. Nevertheless, they were able to select as many features as they felt they learnt without restriction or ranking, and many did not choose grammar, in contrast with the objective results we found.

Interestingly, even a relatively short, five week viewing intervention affected participants' viewing mode preferences and strategies for leisure viewing outside of the classroom. There was a tendency for the participants to switch to watching less with L1 subtitles, and the majority of the participants opted for viewing original version English audio-visual materials without captions rather than with captions. Importantly, the intervention made elementary level students realise that viewing in the original version

is a feasible activity if they have the additional support of L1 subtitles or L2 captions. Finally, viewing strategies were also affected by this shift to uncaptioned audio-visual input. It seems that viewers apply more viewing strategies when they watch with captions.

8.2 Limitations and further research

The empirical results reported in this dissertation should be considered in the light of some limitations. Firstly, convenience samples in the studies may be seen as a limitation. Though it allowed us to collect data in a controlled environment (i.e. we were sure that the participants watched the episodes and did not use any materials during the tests), future research could adopt other means of data collection, especially because viewing television is a common pastime, the participants in future studies could be asked to view the audio-visual materials during their free time (as was done by Vanderplank, 2019). Platforms such as EdPuzzle allow a researcher to oversee an at-home viewing process and make sure the participants complete the episodes (Muñoz et al., under review; Pattemore & Muñoz, 2021).

Another limitation lies in the difficulty level of the TV series. The audio-visual materials may not have been challenging enough for the advanced learners as it appeared that there was a dearth of difficult and low-frequency grammatical constructions for those participants to learn. However, the intact classes in the studies had huge proficiency differences, and the video could not have been more difficult if it was to be comprehensible for the elementary level students, especially in the uncaptioned group. An experiment specifically designed for advanced learners, potentially with another genre of the TV series with higher vocabulary demands (e.g. Suárez et al., 2021) may unveil the mediating factors for higher proficiency students' constructions learning.

In the same line of proficiency level limitations, the low number of participants in the elementary group (14 in total) did not allow us to satisfactorily explore the effects of language learning aptitude on grammar uptake for the elementary and intermediate proficiency groups separately. More studies should focus on elementary, rather than high-intermediate proficiency level learners to support the findings obtained in this dissertation.

An additional limitation lies in the testing materials and the possibility of them being too difficult for the participants. The results of the studies might have been different had we used a simpler recognition test containing, for example, multiple choice questions. Also, as can be seen in Table 4 (Chapter 4), 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. Furthermore, due to the large number of target constructions, the tests did not include distracters which could have limited the testing effect. The already considerable length of the tests and the fact that the target constructions were not presented to the participants (compared to vocabulary studies with meaning recall or recognition tests), led to the decision to not to include any additional test items.

In addition, participants received all the testing materials in written form, which might have disadvantaged the no captions group, who had not seen the constructions in written form in the input. In summary, more research with different types of grammar tests, including auditory[7], oral, and written, is needed to bolster our understanding of the benefits of prolonged exposure to L2 television.

[7] I addressed this limitation in a subsequent research study by implementing an auditory grammaticality judgement test to measure learning of Spanish subjunctive from a TV series with and without captions (Pattemore & Montero Perez, 2022)

One of the shortcomings of Study 2 may be seen in the large number of target constructions 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 for the TE captions group. A follow-up study (Pattemore et al., in progress) addresses this issue and explores reading behaviour of just two episodes of *The Good Place* with enhanced captions, focusing only on five constructions. The eye-tracking data from this study may shed light on whether an excess of enhanced captions was distracting for the viewers. Additionally, future studies could develop a condition where target future constructions are enhanced only the first time they appear in the episode, thus decreasing the frequency of textual enhancement and examining whether this would promote learning. It is important to establish whether textual enhancement could be beneficial in the long-term and what amount of textually enhanced constructions leads to overenhancement.

Another potential limitation is to do with the immediate post-tests. The use of the immediate post-test added some practice that might have enhanced construction learning (although such practice did not provide any feedback, happened only once per target construction, and could not be observed through the statistical analysis). Conversely, increasing the number of immediate post-tests by having one after each episode might have allowed us to measure effects of recency more precisely.

One of the limitations of Study 3 lies in the limited data of students' examples of what they had learnt. Although the questionnaire included this question, it did not elicit many answers. The results may have been clearer had we organised interviews or focus groups after the viewings. Further studies could also include a stimulated recall in order to gain more insights on students' perspective and focus of attention.

Lastly, the results of the questionnaire exploring viewing strategies were insufficient to measure the possible effects of those active measures on extramural language learning. This was mainly because this dissertation focused on learning from the classroom intervention, and not from informal engagement with audio-visual materials. However, this approach allowed us to create a controlled environment that yielded rigorous research output. Future studies could document more closely learners' exposure to videos outside of the classroom, and the strategies that viewers apply through the Experience Sampling Method (see Arndt et al., 2021). The results of using a methodology such as this could shed some light on the effectiveness of viewing strategies in informal settings and provide essential suggestions on how learners can engage with audio-visual input effectively on their own.

8.3 Conclusions and pedagogical implications

This doctoral dissertation provides significant evidence of the potential of audio-visual input on learning L2 grammatical constructions. Collectively, the findings of the three interrelated studies contribute to the increasing area of research on audio-visual input.

Study 1 confirmed that an extensive exposure (227 minutes) to TV series led to significant gains in L2 constructions learning. Furthermore, the group that watched the materials with captions significantly outperformed the uncaptioned group, demonstrating the value of captions for grammar learning. Additionally, it was found that proficiency is one of the mediating factors affecting constructions learning from TV series. The intervention was more valuable for the intermediate group probably because they had reached a sufficient level to benefit from audio-visual materials but still had room to improve in terms of the types of constructions commonly found in the TV series.

Finally, it appears that greater WM capacity and grammar inference ability are required to process audio-visual materials without captions, and that captions in the input can help to compensate for lower levels of the aforementioned cognitive abilities, and allow those students to obtain more value from the audio-visual input.

These findings have several pedagogical implications for both educators and students. Overall, a leisure activity such as viewing L2 television may lead to grammatical constructions uptake, and should be encouraged both in and out of language classes. However, the audio-visual materials should be carefully chosen according to the student's proficiency level. Intermediate and lower-level students may benefit from (captioned) TV series, while higher level students may already be familiar with much of the language used on prime-time TV and may profit the most from other genres of audio-visual materials with a greater volume of lower frequency constructions. Importantly, educators need to be aware that students may differ in WM capacity and language learning aptitude. These are not aspects that are likely to be measured in an educational setting, for reasons both of convenience and egalitarianism, so even in classes of similar proficiencies there will be students with a mixture of cognitive abilities. In order not to disadvantage any students with lower WM capacity or language learning aptitude, teachers should prepare audio-visual materials carefully for the L2 classroom. Captions are generally recommended, as this dissertation has successfully demonstrated that captions themselves are a form of support, and their use could help to level the playing field in language classes where students may have different proficiency levels, WM capacity, and language learning aptitude.

The results of Study 2 also 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. Captions appeared to be more beneficial for long-term retention of the target constructions, while textually enhanced captions resulted in an immediate, short-term, benefit that faded over time.

This 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. This was not unexpected from an intervention that led to incidental learning. A possible way of increasing the learning gains in the classroom may be using audio-visual material in the classroom combined with a focus on form (see for example, Pujadas and Muñoz, 2019). Second, 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 (overenhancement). 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 the arduous process of 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.

Finally, Study 3 also presents novel results for the area of learning from audio-visual input. Specifically, it analysed viewers' feeling of learning from ten episodes of a TV series, and participants' change in viewing preferences in a six-week period. The results support previous research indicating that vocabulary and expressions are perceived to be learnt the most from audio-visual input. Surprisingly, we found no difference in feeling of learning between the different intervention groups. Although previous research has documented that captioned input is more beneficial for content comprehension, vocabulary, and grammar learning (see Montero Perez, 2022), our participants in the captioning groups (both enhanced and unenhanced) did not report more feeling of learning than the no captions group. More research is needed to uncover the reasons behind this incongruity between actual gains and feeling of learning, but such actions as awareness raising and feedback provision are recommended to encourage learners in their progress, and make the learning process through audio-visual input more clear and fruitful.

Importantly, this was the first study to address the variation in feeling of learning from audio-visual input through a proficiency difference perspective. We uncovered that elementary level participants are more likely to perceive language learning from audio-visual input, while the more proficient students need some external assurance that exposure to this type of input can support their language progress significantly. A number of participants expressed the view that audio-visual input is not the optimal way of learning a language and suggested that viewing is not enough to learn from the TV shows. Again, we see the importance of finding ways to raise language learners'

awareness about the learning outcomes (Montero Perez, 2022). Language instructors who use original version audio-visual material educationally should inform their students about how various viewing modes affect learning of different language features and tell them about the importance of viewing the original version media not only for entertainment purposes, but also as a meaningful language activity.

Notably, the third study confirmed that viewing habits and strategies are not stable, and that they can change even over a brief period of time, especially when there is more exposure than usual to L2 input due to a classroom intervention. The results suggest that regular exposure to L2 audio-visual input can boost language confidence to make elementary level students switch from dubbed media to original version with L1 subtitles, and higher proficiency language users are able to watch more without any textual support. The former result contradicts the popular opinion that students may become over reliant on captions, and would not be able to switch to an uncaptioned mode. Although for language development, captions are still recommended, higher proficiency students can be encouraged that they do not need to have captions to enjoy the media. This study provides evidence that viewing with L2 captions is a surmountable stage in becoming a confident L2 viewer.

To conclude, it has been suggested that European countries and regions that practice subtitling rather than dubbing have higher levels of English proficiency (European Commission, 2011; Muñoz & Cadierno, 2021). The demographic population of this doctoral dissertation was Catalan/Spanish bilingual young adults. This group of participants is particularly interesting for this research as Spain is predominantly a dubbing country, and therefore the amount of English programmes in English is limited, with about 69% of American films translated into Spanish (European Commision,

2011). However, the trend to watch in the original version is increasing (Muñoz, 2020) with the growing popularity of on-demand streaming platforms such as Netflix, which reported 4,78 million Spanish registered accounts in 2021 (Statista, 2021).

The importance of the studies like those in this dissertation is that they may promote awareness within language learners and educators about the benefits of original version audio-visual input for language learning, and consequently, motivate learners to view L2 audio-visual input extensively outside the classroom to increase the amount and quality of L2 input they receive.

References

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

Arndt, H. L., Granfeldt, J., & Gullberg, M. (2021). Reviewing the potential of the Experience Sampling Method (ESM) for capturing second language exposure and use. *Second Language Research*. https://doi.org/10.1177/02676583211020055

Avello, D. (in preparation). *Captioned video viewing and L2 learning in primary school students.* (Doctoral dissertation). University of Barcelona.

Ayres, P., & Sweller, J. (2014). The split-attention principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning (Cambridge Handbooks in Psychology)* (p. 206–226). Cambridge University Press. https://doi.org/10.1017/CBO9781139547369.011

Baddeley, A. D. (2012). Working memory: Theories, Models, and Controversies. *Annual Review of Psycology*, 63, 1–29. https://doi.org/10.1146/annurev-psych-120710-100422

Barcelos, A.M.F, & Kalaja, P. (2011). Introduction to beliefs about SLA revisited. *System*, 39(3), 281–289. https://doi.org/10.1016/j.system.2011.07.001

Barón J, & Celaya, M.L. (2022). 'May I do something for you?': The effects of audio-visual material (captioned and non-captioned) on EFL pragmatic learning.

Language Teaching Research, 26(2), 238–255. https://doi.org/10.1177/13621688211067000

Barón, J., & Muñoz, C. (in press). Exploring the effects of audiovisual material and age in pragmatic learning. *RESLA, Spanish Journal of Applied Linguistics*.

Bates D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67, 1–48. https://doi.org/10.18637/jss.v067.i01

Bird, S., Loper, E., & Klein, E. (2009). *Natural Language Processing with Python*. O'Reilly Media Inc.

Buchsbaum, B. R. (2016). Working memory and Language. In G. Hickok, & S.L. Small (Eds.), *Neurobiology of Language* (pp. 863–875). Academic Press. https://doi.org/10.1016/B978-0-12-407794-2.00069-9

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

Carroll, J. B. (1990). Cognitive abilities in foreign language aptitude: Then and now. In T. Parry & C. W. Stansfield (Eds.), *Language aptitude reconsidered* (pp. 11–27). Englewood Cliffs, NJ: Prentice Hall.

Caruana, S. (2021). An Overview of Audiovisual Input as a Means for Foreign Language Acquisition in Different Contexts. *Language and Speech*, 64(4), 1018–1036. https://doi.org/10.1177/0023830920985897

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

Charles, T. J., & Trenkic, D. (2015). Speech segmentation in a second language: The role of bi-modal input. In Y. Gambier, A. Caimi, & C. Mariotti (Eds.), *Subtitles and language learning: Principles, strategies and practical experiences,* (pp. 173–198). Bern: Peter Lang.

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. https://doi.org/10.1080/17501220903388900

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. https://doi.org/10.1080/09571736.2019.1615978

Cobb, T. *Range for texts v.5.1* [computer program]. Accessed 12 Feb 2022 at https://www.lextutor.ca/cgi-bin/range/texts/index.pl

Comeaux, I., & McDonald, J. L. (2018). Determining the effectiveness of visual input enhancement across multiple linguistic cues. *Language Learning*, 68(1), 5–45. https://doi.org/10.1111/lang.12255

Danan, M. (2004). Captioning and subtitling: Undervalued language learning strategies. *Meta: Translators' Journal*, 49(1), 67–77. https://doi.org/10.7202/009021ar

Davies, M. (2008-) *The Corpus of Contemporary American English (COCA): 560 million words, 1990-present.* Available online at https://www.english-corpora.org/coca/. Accessed 1 August 2019.

Davies, M. (2019-) *The TV Corpus: 325 million words, 1950-2018*. Available online at https://www.english-corpora.org/tv/ Accessed 1 August 2019.

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. https://doi.org/10.1075/rcl.00054.kno

De Riso, C. (2021). English audiovisual input: a study of teenagers' habits, attitudes and responses to subtitling versus dubbing. *Rassegna Italiana di Linguistica Applicata*, 1(2), 245–260.

Deslauriers, L., McCarty, L.S., Miller, K., Callaghan, K., & Kestin, G. (2019). Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom. *PNAS*, 116 (39), 19251–19257. https://doi.org/10.1073/pnas.1821936116

De Wilde, V., Brysbaert, M., & Eyckmans, J. (2021). Young learners' L2 English after the onset of instruction: longitudinal development of L2 proficiency and the role of individual differences. *Bilingualism: Language and Cognition*, 24(3), 439–453. https://doi.org/10.1017/S1366728920000747

Diessel, H. (2004). *A dynamic network model of grammatical constructions. In The Acquisition of Complex Sentences* (pp. 13-40). Cambridge University Press. https://doi.org/10.1017/CBO9780511486531.002

Dizon, G. (2021). Subscription video streaming for informal foreign language learning: Japanese EFL students' practices and perceptions. *TESOL Journal*, 12(2), 1–20. https://doi.org/10.1002/tesj.566

Dörnyei, Z. (2003). Attitudes, Orientations, and Motivations in Language Learning: Advances in Theory, Research, and Applications. *Language Learning*, 53(S1), 3–32. https://doi.org/10.1111/1467-9922.53222

Doughty, C. & Williams, J. (eds.) (1998). *Focus on Form in Classroom Second Language Acquisition.* Cambridge University Press.

Dressman, M. (2020). Multimodality and language learning. In M. Dressman, M., Sadler, & W. Randall (Eds.), *The handbook of informal language learning* (pp. 39–55). https://doi.org/10.1002/9781119472384.ch3

d'Ydewalle, G. & Van de Poel, M. (1999). Incidental foreign-language acquisition by children watching subtitled television programs. *Journal of Psycholinguistic Research*, 28(3), 227–244. https://doi.org/10.1023/A:1023202130625

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. https://doi.org/10.1002/9780470756492.ch4

Ellis, N. C. (2006). Language acquisition as rational contingency learning. *Applied Linguistics*, 27(1), 1–24. https://doi.org/10.1093/applin/ami038

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. (2019). Essentials of a Theory of Language Cognition. *The Modern Language Journal*, 103 (S1), 39–60. https://doi.org/10.1111/modl.12532

Ellis, N. C. & Cadierno, T. (2009). Constructing a Second Language Introduction to the Special Section. *Annual Review of Cognitive Linguistics*, 7, 111–139. https://doi.org/10.1075/arcl.7.05ell

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. https://doi.org/10.1111/j.1540-4781.2009.00893.x

Ellis, N. & Ferreira-Junior, F. (2009). Constructions and their acquisition Islands and the distinctiveness of their occupancy. *Annual Review of Cognitive Linguistics*, 7, 187–220. https://doi.org/10.1075/arcl.7.08ell

Ellis, N. C., Römer, U., & O'Donnel, M. (2016). Constructions and Usage-based Approaches to Language Acquisition. *Language Learning*, 66(Supplement 1), 23–44. https://doi.org/10.1111/lang.1_12177

Ellis, R. (2008). Learner beliefs and language learning. *Asian EFL Journal*, 10(4), 7–25.

Ellis, R. (2015). *Understanding Second Language Acquisition*. Oxford University Press.

E-Prime (Version 2.0) [Computer Software]. Sharpsburg, PA: Psychology Software Tools.

European Commission. (2011). *Study on the use of subtitling. The potential of subtitling to encourage foreign language learning and improve the mastery of foreign languages (Report No. EACEA/2009/01)*.

Fernández Guerra, A., & Martínez-Flor, A. (2003). Requests in films and in EFL textbooks: A comparison. *Estudios de lingüística inglesa aplicada*, 4, 17–34.

Field, A. (2018). *Discovering statistics using IBM SPSS Statistics. Fifth edition.* SAGE.

Fievez, I., Montero Perez, M., Cornillie, F., & Desmet, P. (2020). Vocabulary Learning Through Viewing Captioned or Subtitled Videos and the Role of Learner- and Word-Related Factors. *CALICO Journal*, 37(3), 233–253. https://doi.org/10.1558/cj.39370

Fievez, I., Montero Perez, M., Cornillie, F., & Desmet, P. (2021). Promoting incidental vocabulary learning through watching a French Netflix series with glossed captions. *Computer Assisted Language Learning.* https://doi.org/10.1080/09588221.2021.1899244

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., …Zeileis,

A. (2020). *Car: Companion to Applied Regression // R package version 3.0-10.* 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). De Gruyter Mouton. https://doi.org/10.1515/9783110363708-005

Frumuselu, A. D., De Maeyerb, S., Doncheb, V., Gutiérrez-Colon Plana, M. Del M. (2015). Television series inside the EFL classroom: Bridging the gap between teaching and learning informal language through subtitles. *Linguistics and Education*, 32(B), 107–117. https://doi.org/10.1016/j.linged.2015.10.001

García-Amaya, L., & Cintrón-Valentín, M. (2021). The Effects of Textually Enhanced Captions on Written Elicited Imitation in L2 Grammar. *The Modern Language Journal*, 105(4), 919–935. https://doi.org/10.1111/modl.12740

Gass, S. M., Spinner, P., & Behney, J. (2018). *Salience in second language acquisition*. Routledge. https://doi.org/10.4324/9781315399027

Gass, S., Winke, P., Isbell, D. R., & Ahn, J. (2019). How captions help people learn languages: A working-memory, eye-tracking study. *Language Learning & Technology*, 23(2), 84–104.

Gesa, F., & Miralpeix, I. (2022). Effects of watching subtitled TV series on foreign language vocabulary learning: does learners' proficiency level matter? In C. Lütge (Ed.), *Foreign Language Learning in the Digital Age. Theory and Pedagogy for*

Developing Literacies (pp. 159–173). Routledge. https://doi.org/10.4324/9781003032083-14

Gilabert, R., & Muñoz, C. (2010). Differences in Attainment and Performance in a Foreign Language: The Role of Working Memory Capacity. *International Journal of English Studies*,10(1), 19–42. https://doi.org/10.6018/ijes/2010/1/113961

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

Goldberg, A. E., & Casenhiser, D. (2006b). English Constructions. In April McMahon and Bas Aarts (Eds.), *The Handbook of English Linguistics* (pp. 343–355). Blackwell Publishers. https://doi.org/10.1002/9780470753002.ch15

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). 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. https://doi.org/10.1111/j.0023-8333.2005.00295.x

Granena, G. (2013). Cognitive aptitudes for second language learning and the LLAMA Language Aptitude Test. *Language Learning & Language Teaching*, 35, 105–130. https://doi.org/10.1075/lllt.35.04gra

Granena, G. & Long, M (Eds.). (2013). *Sensitive periods, language aptitude, and ultimate attainment.* Philadelphia, PA: John Benjamins. https://doi.org/10.1075/lllt.35

Gries, S., & Wulff, S. (2009). Psycholinguistic and corpus-linguistic evidence for L2 constructions. *Annual Review of Cognitive Linguistics*, 7, 163–186. https://doi.org/10.1075/arcl.7.07gri

Han, Z. Park, E., & Combs, C. (2008). Textual Enhancement of Input: Issues and Possibilities. *Applied Linguistics*, 29(4), 597–618. https://doi.org/10.1093/applin/amn010

Handbreak [Computer software] (n.d.). Retrieved from https://handbrake.fr/

Hede, A. (2002). An integrated model of multimedia effects on learning. *Journal of Educational Multimedia and Hypermedia*, 11(2), 177–191.

Hummel, K. (2009). Aptitude, phonological memory, and second language proficiency in novice adult learners. *Applied Psycholinguistics*, 30, 225–249. https://doi.org/10.1017/S0142716409090109

Indrarathne, B., Ratajczak, M., Kormos, J. (2018). Modelling Changes in the Cognitive Processing of Grammar in Implicit and Explicit Learning Conditions: Insights From an Eye-Tracking Study. *Language Learning*, 68(3), 669–708. https://doi.org/10.1111/lang.12290

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. https://doi.org/10.1017/S027226311800013X

Khazdouzian, Y., Celaya, M.L., Barón, J. (2021). When watching is not enough: The effects of captions on L2 pragmatics acquisition and awareness. *RAEL - Revista Electrónica de Lingüística Aplicada*, 19, 90–107.

Knight, B. (2018). *How long does it take to learn a language?* CUP.

Kormos, J. (2013). New conceptualization of aptitude. In G. Gisela & M. Long (Eds.), *Sensitive periods, language aptitude, and ultimate attainment* (pp. 131–152). Philadelphia, PA: John Benjamins. https://doi.org/10.1075/lllt.35.05kor

Krashen, S.D. (1985). *The input hypothesis: Issues and implications.* Addison-Wesley Longman Ltd.

Kuppens, A. H. (2010). Incidental foreign language acquisition from media exposure. *Learning, Media and Technology*, 35(1), 65–85. https://doi.org/10.1080/17439880903561876

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. https://doi.org/10.4000/asp.3104

Lai, C. (2019). Learning beliefs and autonomous language learning with technology beyond the classroom. *Language Awareness*, 28(4), 291–309. https://doi.org/10.1080/09658416.2019.1675679

Laufer, B. (1989). What percentage of text lexis is essential for comprehension? In C. Lauren & M. Nordman (Eds.), *Special language: From humans thinking to thinking machines* (pp. 316–323). Clevedon, UK: Multilingual Matters

Lee, M. & Révész, A. (2018). Promoting grammatical development through textually enhanced captions: an eye-tracking study. *The Modern Language Journal*, 102(3), 557–577. https://doi.org/10.1111/modl.12503

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

Lee, M., & Revesz, A. (2021). The Role of Working Memory in Attentional Allocation and Grammatical Development under Textually-enhanced, Unenhanced and No Captioning Conditions. *Journal for the psychology of language learning*, 3(1), 6–25. https://doi.org/10.52598/jpll/3/1/1

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. https://doi.org/10.1017/S0272263108080479

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*. 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). Routledge. https://doi.org/10.4324/9781315399027-9

Li, S. (2016). The construct validity of language aptitude. *Studies in Second Language Acquisition*, 38(4), 801–842. https://doi.org/10.1017/S027226311500042X

Liao, S., Kruger, J. L., & Doherty, S. (2020). The impact of monolingual and bilingual subtitles on visual attention, cognitive load, and comprehension. *The Journal of Specialised Translation*, 33, 70–98.

Lin, P. (2021). Developing an intelligent tool for computer-assisted formulaic language learning from YouTube videos. *ReCALL*, 1–16. https://doi.org/10.1017/S0958344021000252

Linck, J.A., Osthus, A., Koeth, J.T., & Bunting M.F. (2014). Working memory and second language comprehension and productions: A meta-analysis. *Psychonomic Bulletin & Review*, 21, 861–883. https://doi.org/10.3758/s13423-013-0565-2

Lindgren, E., & Muñoz, C. (2013). The influence of exposure, parents, and linguistic distance on young European learners' foreign language comprehension. *International Journal of Multilingualism*, 10(1), 105–129. https://doi.org/10.1080/14790718.2012.679275

Long, M. (2020). Optimal input for language learning: Genuine, simplified, elaborated, or modified elaborated? *Language Teaching*, 53, 169–182. https://doi.org/10.1017/S0261444819000466

Madlener, K. (2015). *Frequency effects in instructed second language acquisition*. De Gruyter Mouton. https://doi.org/10.1515/9783110405538

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

Mariotti, C. (2014). A survey on stakeholders' perceptions of subtitles as a means to promote foreign language learning. In Y. Gambier, A. Caimi, C. Mariotti (Eds.), *Subtitles and Language Learning* (pp. 83–103). Peter Lang.

Markham, P.L. (1989). The effects of captioned videotapes on the listening comprehension of beginning, intermediate, and advanced ESL students. *Educational Technology*, 29(10), 38–41.

Martin, K. & Ellis, N. (2012). The roles of phonological short-term memory and working memory in L2 grammar and vocabulary learning. *Studies in Second Language Acquisition*, 34 (3), 379–413. https://doi.org/10.1017/S0272263112000125

Martínez-Flor A. (2008) Analysing Request Modification Devices in Films: Implications for Pragmatic Learning in Instructed Foreign Language Contexts. In:

Soler E.A., Jordà M.S. (Eds.) *Intercultural Language Use and Language Learning* (pp. 245–280).. Springer. https://doi.org/10.1007/978-1-4020-5639-0_13

Matielo, R., D'Ely, R., & Baretta, L. (2015). The effects of interlingual and intralingual subtitles on second language learning/acquisition: a state-of-the-art review. *Trabalhos em Linguística Aplicada*. 54, 161–182. https://doi.org/10.1590/0103-18134456147091

Matielo, R., Pires de Oliveira, R., & Baretta, L. (2018). Subtitling, worling memory, and L2 learning: A correlational study. *Revista Brasileira de Linguística Aplicada*, 18(3), 665–696. https://doi.org/10.1590/1984-6398201812773

Mayer, R. E. (2009). *Multimedia learning (2nd ed.).* New York: Cambridge University Press.

https://doi.org/10.1017/CBO9780511811678

Mayer, R. (Ed.). (2014). *The Cambridge Handbook of Multimedia Learning (Cambridge Handbooks in Psychology).* Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139547369

Mayer, R.E., Fiorella, L. & Stull, A. (2020). Five ways to increase the effectiveness of instructional video. *Educational Technology Research and Development*, 68, 837–852. https://doi.org/10.1007/s11423-020-09749-6

Meara, P.M., & Milton, J. (2003). *X_Lex: The Swansea Vocabulary Levels Test.* Newbury: Express Publishing.

Meara, P.M. (2005). *Llama Language Aptitude Tests*. Swansea: Lognostics.

Meara, P.M., & Miralpeix, I. (2017). *Tools for Researching Vocabulary*. Bristol: Multilingual Matters.

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

Milton, J., & Alexiou, T. (2009) Vocabulary Size and the Common European Framework of Reference for Languages. In: B. Richards B, M.H. Daller, D.D. Malvern , P. Meara, J. Milton, & J. Treffers-Daller (Eds.), *Vocabulary Studies in First and Second Language Acquisition* (pp. 194–211). Palgrave Macmillan: London. https://doi.org/10.1057/9780230242258_12

Mitterer, H., & McQueen, J.M. (2009). Foreign Subtitles Help but Native-Language Subtitles Harm Foreign Speech Perception. *PLoS ONE*, 4(11), 77–85. https://doi.org/10.1371/journal.pone.0007785

Mirzaei, M., Meshgi, K., Akita, Y., & Kawahara, T. (2017). Partial and synchronized captioning: A new tool to assist learners in developing second language listening skill. *ReCALL*, 29(2), 178–199. https://doi.org/10.1017/S0958344017000039

Miyake, A., & Friedman, N. P. (1998). Individual differences in second language proficiency: Working memory as language aptitude. In A. Healy & L. Bourne (Eds.), *Foreign language learning: Psycholinguistics studies on training and retention* (pp.339–364). Mahwah, NJ: Lawrence Erlbaum.

Montero Perez. M. (2020). Incidental vocabulary learning through viewing video: The role of vocabulary knowledge and working memory. *Studies in Second Language Acquisition*, 1–25. https://doi.org/10.1017/S0272263119000706

Montero Perez, M. (2022). Second or foreign language learning through watching audio-visual input and the role of on-screen text. *Language Teaching*, 1–30. https://doi.org/10.1017/S0261444821000501

Montero Perez, M., Peters, E., & Desmet, P. (2014). Is less more? Effectiveness and perceived usefulness of keywords and full captioned video for L2 listening comprehension. *ReCALL*, 26(1), 21–43. https://doi.org/10.1017/S0958344013000256

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. Retrieved from http://llt.msu.edu/issues/february2014/monteroperezetal.pdf

Montero Perez, M., & Rodgers, M.P.H. (2019). Video and language learning. *The Language Learning Journal*, 47 (4), 403–406. https://doi.org/10.1080/09571736.2019.1629099

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. (2020). Boys like games and girls like movies. Age and gender differences in out-of-school contact with English. *RESLA, Spanish Journal of Applied Linguistics*, 33 (1), 171–201. https://doi.org/10.1075/resla.18042.mun

Muñoz, C & Cadierno, T. (2021) How do differences in exposure affect English language learning? A comparison of teenagers in two learning environments. *Studies in Second Language Learning and Teaching*, 11, 185–212. https://doi.org/10.14746/ssllt.2021.11.2.2

Muñoz, C., Pattemore, A., & Avello, D. (under review). Exploring repeated captioning viewing as a way to promote vocabulary learning: Time lag between repetitions and learner factors.

Muñoz, C., & Peters, E. (2020). Special issue: Language learning from multimodal input. (2020). *Studies in Second Language Acquisition*, 42(3). http://doi.org/10.1017/S0272263120000340

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

Nation, P. (2007). The four strands. *International Journal of Innovation in Language Learning and Teaching*, 1(1), 1–12. https://doi.org/10.2167/illt039.0

Nation, P. (2015). Principles guiding vocabulary learning through extensive reading. *Reading in a Foreign Language*, 27 (1), 136–145.

Nguyen, C., & Boers, F. (2018). The effect of content retelling on vocabulary uptake from a TED talk. *TESOL Quarterly*, 53(1), 5–29. https://doi.org/10.1002/tesq.441

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

Oxford, R.L. (1989). Use of language learning strategies: A synthesis of studies with implications for strategy training. *System*, 17(2), 235–247. https://doi.org/10.1016/0346-251X(89)90036-5

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

Pattemore, A., & Muñoz, C. (2021, August 26-27). The role of time distribution of captioned TV series on learning L2 constructions [Paper presentation]. EuroCALL 2021, Paris, France.

Pattemore, A., & Montero Perez, M. (2022, April 27–29). Learning of Spanish subjunctive through original version audio-visual input [Paper presentation]. AESLA 39, Las Palmas de Gran Canaria, Spain.

Pattemore, A., Muñoz, C., & Suárez, M.M. (in progress). Processing of multiple textually enhanced grammatical constructions in captioned audio-visual input: An eye-tracking study.

Pellicer-Sánchez, A. (2017). Learning L2 collocations incidentally from reading. *Language Teaching Research*, 21(3), 381–402. https://doi.org/10.1177/1362168815618428

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. Retrieved from: https://www.cambridge.org/partnership/research/impact-usage-based-approaches-second-language-learning-and-teaching

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

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

Plonsky, L., & Oswald, F.L. (2014). How Big is "Big"? Interpreting Effect Sizes in L2 Research. *Language Learning*, 64(4), 878–912. https://doi.org/10.1111/lang.12079

Puimège, E., & Peters, E. (2020). Learning formulaic sequences through viewing L2 television and factors that affect learning. *Studies in Second Language Acquisition*, 42(3), 525–549. http://doi.org/10.1017/S027226311900055X

Puimège, E., Montero Perez, M., & Peters, E. (2021). 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* (Unpublished 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), 551–575. https://doi.org/10.1017/S0272263120000042

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

Robinson, P. (2012). Individual differences, aptitude complexes, SLA processes and aptitude test development. In M. Pawlak (Ed.), *New perspectives on individual differences in language learning and teaching* (pp. 57–76). Oxford: Springer. https://doi.org/10.1007/978-3-642-20850-8_4

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). Routledge.

Rodgers, M. P. H. (2013). *English language learning through viewing television: An investigation of comprehension, incidental vocabulary acquisition, lexical*

coverage, attitudes, and captions (Unpublished doctoral dissertation). Victoria University of Wellington.

Rodgers, M.P.H. (2016). Extensive listening and viewing: The benefits of audiobooks and television. *The European Journal of Applied Linguistics and TEFL*, 5(2), 43–57.

Rodgers, P.H., & Webb, S. (2017). The effects of captions on EFL Learners' comprehension of English-language television programs. *CALICO Journal*, 34(1), 20–38. https://doi.org/10.1558/cj.29522

Rodgers, P.H., & Webb, S. (2020). Incidental vocabulary learning through viewing television. *ITL-International Journal of Applied Linguistics*, 171(2), 191–220. https://doi.org/10.1075/itl.18034.rod

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. https://doi.org/10.1075/ijlcr.17015.rom

Sagarra, N. (2017). Longitudinal effects of working memory on L2 grammar and reading abilities. *Second Language Research*, 33(3), 341–363. https://doi.org/10.1177/0267658317690577

Salomon, G. (1984). Television is "easy" and print is "tough": The differential investment of mental effort in learning as a function of perceptions and attributions. *Journal of Educational Psychology*, 76(4), 647–658. https://doi.org/10.1037/0022-0663.76.4.647

Sergeeva, M. (2021). *Subtitled Audiovisual Materials in the FL Classroom: Teachers' Perspective* (Unpublished Master's thesis). University of Barcelona.

Sharwood Smith, M. (1993). Input enhancement in instructed SLA. *Studies in Second Language Acquisition*, 15, 165–179. https://doi.org/10.1017/S0272263100011943

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

Skehan, P. (2002). Theorising and updating. In P. Robinson (Ed.), *Individual differences and instructed language learning* (pp. 69–93). Philadelphia, PA: John Benjamins. https://doi.org/10.1075/lllt.2.06ske

Sockett, G. & Kusyk, M. (2015). Online informal learning of English: frequency effects in the uptake of chunks of language from participation in web-based activities. In T. Cadierno, & S.W. Eskildsen (Eds.), *Usage-based perspectives on second language acquisition* (pp. 153–177). Berlin: De Gruyter Mouton. https://doi.org/10.1515/9783110378528-009

Statista (2021). *Number of Netflix subscribers in selected countries in Europe as of 2nd quarter 2021*. Retrieved from https://www.statista.com/statistics/448394/number-of-netflix-subscribers-europe-by-country/

Stewart, M., & Pertusa, I. (2004). Gains to Language Learners from Viewing Target Language Closed-Captioned Films. *Foreign Language Annals*, 37(3), 438–442. https://doi.org/10.1111/j.1944-9720.2004.tb02701.x

Suárez, M.M. (2010). *Language Aptitude in Young Learners: The Elementary Modern Language Aptitude Test in Spanish and Catalan* (Unpublished doctoral dissertation). Universitat de Barcelona. http://hdl.handle.net/10803/38244

Suárez, 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

Suárez, M, Gilabert, R., & Moskvina, N. (2021). The mediating role of vocabulary size, working memory, attention and inhibition in early vocabulary learning under different TV genres: An exploratory study. *TESOL Journal*, 12(4). https://doi.org/10.1002/tesj.637

Sydorenko, T. (2010). Modality of input and vocabulary acquisition. *Language Learning & Technology*, 14(2), 50–73.

Swain, M. (1985). Communicative Competence: Some Roles of Comprehensible Input and Comprehensible Output in its Development. In S. Gass & C. Madden (Eds.), *Input in Second Language Acquisition* (pp. 235–253). Newbury House.

Sweller, J. (2005). Implications of Cognitive Load Theory for Multimedia Learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (p. 19–30). Cambridge University Press. https://doi.org/10.1017/CBO9780511816819.003

Tagarelli, K. M., Borges Mota, M., & Rebuschat, P. (2015). Working memory, learning conditions, and the acquisition of L2 syntax. In Z. Wen, M. Borges Mota, & A.

McNeill (Eds.), *Working memory in second language acquisition and processing: theory, research and commentary* (pp. 224–247). Bristol: Multilingual Matters. https://doi.org/10.21832/9781783093595-017

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. https://doi.org/10.1177/1362168807080962

Tang, M., & Tian, J. (2015). Associations between Chinese EFL graduate students' beliefs and language learning strategies. *International Journal of Bilingual Education and Bilingualism*, 18(2), 131–152. https://doi:10.1080/13670050.2014.882882

Taylor, G. (2005). Perceived processing strategies of students watching captioned video. *Foreign Language Annals*, 38(3), 422–427. https://doi.org/10.1111/j.1944-9720.2005.tb02228.x

Teng, M.F. (2019). Maximizing the potential of captions for primary school ESL students' comprehension of English-language videos. *Computer Assisted Language Learning*, 32(7), 665–691. https://doi.org/10.1080/09588221.2018.1532912

Teng, M. F. (2021). *Language Learning Through Captioned Videos: Incidental Vocabulary Acquisition.* Routledge. https://doi.org/10.4324/9780429264740

Tomasello, M. (2005). *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press. https://doi.org/10.2307/j.ctv26070v8

Tomasello, M. (2009). The usage-based theory of language acquisition. In E. Bavin (Ed.), *The Cambridge Handbook of Child Language (Cambridge Handbooks in Language and Linguistics* (pp. 69–88). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511576164.005

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

Usó-Juan, E., & Martínez-Flor, A. (2021). Foresting learner's (meta)pragmatic awareness through film analysis. *Language Value*, 14(1), 85–111. https://doi.org/10.6035/languagev.5821

Vanderplank, R. (1988), The value of teletext sub-titles in language learning. *ELT Journal*, 42(4), 272–281. https://doi.org/10.1093/elt/42.4.272

Vanderplank, R. (2010). Déjà vu? A decade of research on language laboratories, television and video in language learning. *Language Teaching*, 43(1), 1–37. http://doi.org/10.1017/S0261444809990267

Vanderplank, R. (2016). *Captioned media in foreign language learning and teaching: Subtitles for the deaf and hard-of-hearing as tools for language learning*. Oxford: Palgrave Macmillan. https://doi.org/10.1057/978-1-137-50045-8

Vanderplank, R. (2019). 'Gist watching can only take you so far': attitudes, strategies and changes in behaviour in watching films with captions, *The Language Learning Journal*, 47(4), 407–423. https://doi.org/10.1080/09571736.2019.1610033

Vanderplank, R. (2020). Video and Informal Language Learning. In M. Dressman, M., Sadler, & W. Randall (Eds.), *The handbook of informal language learning* (pp.183–201). https://doi.org/10.1002/9781119472384.ch12

Van Lommel, S., Laenen, A., & D'Ydewalle, G. (2006). Foreign-grammar acquisition while watching subtitled television programmes. *British Journal of Educational Psychology*, 76, 243–258. https://doi.org/10.1348/000709905X38946

Wang, A., & Pellicer-Sánchez, A. (2022). Incidental vocabulary learning from bilingual subtitled viewing: An eye-tracking study. *Language Learning*. https://doi.org/10.1111/lang.12495

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). 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. https://doi.org/10.1017/S0272263114000606

Webb, S., & Rodgers, M.P.H. (2009). Vocabulary Demands of Television Programs. *Language Learning*, 59(2), 335–366. https://doi.org/10.1111/j.1467-9922.2009.00509.x

Wen, Z. (2019). Working memory as language aptitude. In Z. Wen, P. Skehan, A. Biedroń, S. Li, & R.L. Sparks (Eds.), *Language Aptitude: Advancing Theory, Testing, Research and Practice* (pp. 187–214). Routledge. https://doi.org/10.4324/9781315122021

Wen, Z., Biedrón, A. & Skehan, P. (2016). Foreign language aptitude theory: Yesterday, today and tomorrow. *Language Teaching*, 50(1), 1–31. https://doi.org/10.1017/S0261444816000276

Wiley, J., Sanchez, C.A., & Jaeger, A.J. (2014). The individual differences in working memory capacity principle in multimedia learning. In R. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning (2nd ed.)* (pp. 598–619). New York: Cambridge University Press. https://doi.org/10.1017/CBO9781139547369.029

Wilkinson, D., & Apic, O. (2018). *Language Reactor* [Computer Software]. Retrieved from https://www.languagereactor.com/

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

Winke, P. (2013). An investigation into second language aptitude for advanced Chinese language learners. *The Modern Language Journal*, 97(1), 109–130. https://doi.org/10.1111/j.1540-4781.2013.01428.x

Winke P., Gass S., & Sydorenko T. (2010). The Effects of Captioning Videos Used for Foreign Language Listening Activities. *Language Learning & Technology*, 14 (1), 65–86.

Wisniewska, N., & Mora, J. C. (2020). Can captioned video benefit second language pronunciation? *Studies in Second Language Acquisition*, 42(3), 599–624. http://doi.org/10.1017/S0272263120000029

Xu, S., Fang, X., Brzezinski, J., & Chan, S. (2008). Development of a Dual-Modal Presentation of Texts for Small Screens. *International Journal of Human-Computer Interaction*, 24(8), 1–18. https://doi.org/10.1080/10447310802537566

Yalçın, S., Çeçen, S. & Erçetin, G. (2016). The relationship between aptitude and working memory: an instructed SLA context. *Language Awareness*, 25 (1-2), 144–158. https://doi.org/10.1080/09658416.2015.1122026

Appendices

Appendix A. Pre-course questionnaire

Dear students,
We would be grateful if you could answer the following questions about your contact with English outside of the classroom. All the information provided will be treated confidentially and will be only used for research purposes.
Thank you.

1. Surname and name

2. How old are you?

3. Indicate how often you do the following activities in **English**.

	Never	Between 1-3 times a month	Between 1-3 times a week	Between 4-6 times a week	Every day
Playing video games					
Watching movies and/or TV series					
Listening to music					
Reading (e.g. books, magazines, articles, comic books)					
Surfing Internet					
Writing (e.g. e-mail, chat, WhatsApp, Facebook, Instagram, Twitter)					
Watching YouTube videos					
Speaking English with friends					
Speaking English with relatives					
Speaking English with tourists					
Speaking English at work					

3. Have you watched "*The Good Place*" TV series?

☐ Yes

☐ No

4. Do you watch films and/or TV series in English outside of the classroom (e.g. at home, at the cinema)?

☐ Yes

☐ No*

**If you answered "no" you do not have to answer the rest of the questions.*

5. What films and/or TV series have you watched in English recently? Write down at least 3.

☐

6. Do you watch films and/or TV series with subtitles? If yes, specify the language of subtitles.

☐ With Catalan/Spanish subtitles

☐ With English subtitles

☐ Without subtitles

☐ Other: ☐

7. Indicate how often you watch films and/or TV series in English.

☐ 2–3 times a month

☐ Once a week

☐ Between 1–3 times a week

☐ Between 4–6 times a week

☐ Every day

8. When you watch films and/or TV series in English, approximately how long do you watch for?

☐ Less than 20 minutes

☐ Between 20–30 minutes

☐ Between 30 minutes–1 hour

☐ Between 1–3 hours

☐ More than 3 hours

9. When I watch films and/or TV series in English with subtitles…

	Never	Sometimes	Often	Very often	Always
I try to listen to the audio before reading the subtitles					
I read the subtitles before listening to the audio					
I only read the subtitles if I do not understand the audio					

10. Why do you choose to watch films and/or TV series in English?

☐ Because I want to watch the most recent content before it is translated

☐ Because I like to watch in the original version

☐ Because there is no dubbed (Catalan/Spanish) version

☐ Because I want to improve my English

☐ Other: ☐

11. Do you do any of the following to improve your English while watching films and/or TV series in English?

☐ I replay scenes

☐ I stop the video to write down new vocabulary and expressions

☐ I look up unknown vocabulary and expressions in the dictionary

☐ I try to use the vocabulary and expressions from the video

☐ I pay attention to new words and expressions

☐ I do nothing

☐ Other: ☐

12. Do you feel that you are learning new vocabulary / expressions / pronunciation / grammar from the films and/or TV series you watch in English?

☐ Yes, vocabulary

☐ Yes, expressions

☐ Yes, pronunciation

☐ Yes, grammar

☐ I do not know if I am learning

☐ Other: ☐

13. Specify what you have learnt from films and/or TV series in English (e.g. new words, expressions, pronunciation, grammar). Write down at least 3 examples.

☐

Thank you for your answers.

Appendix B. The n-gram Python script

This script was used to identify the most frequently occurring n-grams in the chosen TV series[8].

```
        import string
import unicodedata
from nltk.tokenize import word_tokenize
from nltk.util import ngrams

def get_ngrams(text, n):
    n_grams = ngrams(text, n)
    return [ ' '.join(grams) for grams in n_grams]

#Open a file
success = False
while success == False:
    try:
        fname = input("Please enter the file name: ")
        fhand = open(fname, encoding="utf8")
        success = True
    except:
        print("File not found.", fname)

#Find out N-gram length
success2 = False
while success2 == False:
    try:
        ngramlength = int(input("What length n-grams? "))
        if ngramlength > 10:
            print("please use a slightly lower number (less than 11)")
            continue
        success2 = True
    except:
        print("Please enter n-gram length using digits")

wordlist = list()
gramlist = list()
counts = dict()

#strip punctuation and line breaks.
for line in fhand:
    line = line.rstrip()
```

[8] I am grateful to Matthew Pattemore for assisting with the script and generously allowing me to include it as an appendix.

```
    line = line.translate(line.maketrans("", "", string.punctuation))
    line = line.lower()
    words = line.split()
    for word in words:
        wordlist.append(word)

#update the appendix dictionary with the n-grams
gramlist.append(get_ngrams(wordlist, ngramlength))
for listo in gramlist:
    for phrase in listo:
        if phrase not in counts:
            counts[phrase] = 1
        else:
            counts[phrase] += 1

# Sort the dictionary by value
lst = list()
for key, val in list(counts.items()):
    lst.append((val, key))
lst.sort(reverse=True)

#Work out how many n-grams the user wants
success3 = False
while success3 == False:
    try:
        outputnumber = int(input("How many n-grams do you want?"))
        success3 = True
    except:
        print("Please enter number of n-grams using digits")

#Print the user's chosen number of n-grams
for key, val in lst[:outputnumber]:
    print(key, val)
```

Appendix C. Pre-/delayed post-test test items

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

1. The young man was disappointed because his best friend had never betrayed him before.

   LET The young man felt _____because his best friend had never betrayed him before.

2. You should have met her yesterday.

   SUPPOSED You_____her yesterday.

3. What is your job?

   FOR What do you _____?

4. Alice has prepared the dinner tonight.

   BY The dinner _____ Alice tonight.

5. Can he borrow your car, please?

   LET Will you _____ your car?

6. I really liked that lesson.

   LIKE I _____ that lesson.

7. I will not go to another bar, I would like to go home.

   JUST I _____ go home.

8. I ate too much ice cream and now I don't feel well.

   WISH I _____ so much ice cream.

9. Tina wants to eat out, not at home.

   RATHER Tina _____eat out.

10. "We can help you with finding a flat," said my friends.

    HELP My friends said _____ with finding a flat.

11. It would be a good idea to go to your place.

    WHY _____ to your place?

12. Person A: She totally hates it when people tell her what to do.

    HATE Person B: I agree with you, she _____it when people tell her what to do.

13. "You should prepare for the test next week," said the teacher.

    WANTS The teacher _____ for the test next week.

14. The only reason I did it was to help you.

    JUST I _____ help you.

15. Could you bring my book, please?

    NEED I _____ my book, please.

16. It is not important at all.

    BIG It is _____ .

17. I totally understand you.

    NO Say _____.

18. It is prohibited to smoke here.

    ARE You _____ smoke here.

19. It takes most people some time to understand how the new software works.

    FIGURE It takes most people some time to _____ new software.

20. “I will buy you a present,” my mum said.

    BUY My mum said _____ a present.

21. I prefer to stay at the library.

    RATHER I _____ at the library.

22. I want to go on holiday, but I haven’t got enough money.

    WISH I _____ enough money to go on holiday.

23. They are very serious about this relationship.

    BIG This relationship _____ for them.

24. “You can go to the beach,” my parents said.

    LET My parents _____ to the beach.

25. All of the employees should attend today's meeting.

    SUPPOSED All of the

    employees_____today's meeting.

26. John is making some coffee for us.

    IS The coffee _____ by John.

27. He always went out with his friends when he was younger.

    TO He _____with his friends when he

    was younger.

28. I got the point of your conversation.

    NO Say _____.

29. I think you should apply for this scholarship.

    WHY _____ for this

    scholarship?

30. You can take a friend with you to the ceremony.

    ARE You _____ a friend with you

    to the ceremony.

II. Complete the gaps with the appropriate word or contraction (e.g. is, don't)

31. Anne is thirsty, she says to her friend, " _____ buy some water".

32. Colin doesn't know how to drive. Jane doesn't,_____.

33. I am counting on you. Please don't let me_____.

34. When I told my mum I had a new boyfriend, the first question she asked was,

    "What does he do for a _____?"

35. If I make a promise, I never_____ it.

36. James wants to go to the cinema. He says to his girlfriend, "

_____go to the cinema".

37. He is not a part of this group, he doesn't_____here.

38. John doesn't like coffee, and I don't, _____.

39. A: "Sharon looks different now".

B : "Yes, she_____ to have long dark hair, now she doesn't."

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

41. We had to _____ out the connection between two events.

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

43. Soulmates are people who _____together.

III. Complete the sentences using a form of the words in brackets

44. This time next week I _____ (lie)

on the beach.

45. What are the different types of_____(cactus) in Mexico?

46. Mr. Jones has a mouse problem in his house, he has found about

50_____ (mouse) in his basement.

47. Mr Simon was _____ (resign),

but the manager offered him a better salary so he didn´t resign.

48. _____(cold) it got, _____(many)

clothes they had to put on to keep warm.

49. This time tomorrow he _____ (watch) a new episode of *Game of Thrones*.
50. The fisherman has sold about 500_____(shrimp) this morning.
51. A lot of _____ (person) use the underground every day.
52. Sorry, I'm late. I didn't realise the meeting _____(will) go on all day.
53. _____ (old) he gets, _____ (tall) he grows.
54. You can see a lot of _____(cactus) in the desert.

Appendix D. Immediate post-test test items

The partial immediate post-tests were completed after viewing the second episode of the week. There were five partial immediate post-tests.

I. Partial immediate post-test 1 after viewing episode 2

*Complete each sentence with **two to five words**, including the word in bold*

1. 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.

2. Sue is washing the car.

   IS The car _____ by Sue.

3. I felt like I betrayed my parents when I failed my exam.

   LET I felt like I _____when I failed my exam.

4. My father never talked about his job.

   FOR My father never talked about what

   _____.

5. John has baked the pie.

   BY The pie _____John.

6. You are coming to the party – you won't disappoint me, right?

   LET You are coming to the party – you won´t

   _____, right?

Complete the gaps with the appropriate word or contraction (e.g. is, don't)

7. John wants to play tennis with Mary, he says to her:_____ play tennis.

8. Kate is hungry, she says to her boyfriend: _____ go to a restaurant.

Complete the sentences using the words in brackets

9. How many _____ (mouse) are there in the pet shop?

10. There are a lot of _____ (shrimp) in the Mediterranean sea.

II. Partial immediate post-test 2 after viewing episode 4

*Complete each sentence with **two to five words**, including the word in bold.*

1. Could you lend me your dress, please?

   LEND I want _____ your dress, please.

2. A: I can't believe that Paul has a new girlfriend, I still love him!

   HAVE B: You have to forget about him, he _____ a new girlfriend.

3. I think she is lost, she shouldn´t be here.

   HERE I think she is lost, she doesn´t _____.

4. I'm sorry I didn't come to your party, I had a terrible headache and my only wish was to go to bed.

   JUST I had a terrible headache and _____to bed.

5. A: You didn't wash the dishes when I asked you to.

   WASH B: That's not true! I _____the dishes!

6. "You need to finish the report as soon as possible," said my boss.

   ME My boss wants _____the report as soon as possible.

7. Their relationship is an example of true love: They are meant for each other.

   TOGETHER They_____.

8. Did she like the movie?

   LIKED She _____, _____ she?

9. The only reason I did it was to help you.

JUST I _____help you.

Complete the gap with one appropriate word.

10. Let's go out tonight, _____ we?

III. Partial immediate post-test 3 after viewing episode 6

*Complete each sentence with **two to five words**, including the word in **bold***

1. I lost the pen I borrowed from her, but she said it wasn't a problem.

BIG I lost the pen I borrowed from her, but she said it was

_____.

2. I totally understand you.

NO Say_____.

3. "We can drive you home," said my parents.

DRIVE My parents said

_____home.

4. You can enter the night club.

ARE You_____the night club.

5. "I will take you to the movies," said John.

TAKE John said_____to the

movies.

6. Why do you think it is so important?

BIG Why do you think it is such _____?

7. Kate forbids her dog to sleep on the bed.

NOT Her dog is_____sleep on the

bed.

Complete the sentences using a form of the words in brackets

8. By this time next month I

_____ (work) at my dream job.

9. At this time tomorrow I _____ (eat)

a delicious gelato.

Complete the gaps with one word only

10. Try to _____ out a way to help her.

11. Politicians lie to people, they make promises before elections, but then _____

their promises after they've won.

IV. Partial immediate post-test 4 after viewing episode 8

Complete the sentences using a form of the words in brackets

1. You are home early. I thought you_____(will) play tennis with Alex all afternoon long.
2. _____(hard) he works,
   _____(successful) he becomes.
3. If you go to the botanical garden you will see lots of different
   _____(cactus).
4. _____ (young) you are, _____(easy) you find it to learn things.
5. The ocean near the Maldive islands is full of _____ (octopus).

*Complete each sentence with **two to five words**, including the word in **bold***

6. I travelled a lot when I was younger.

TO I _____ a lot when I was younger.

7. We expect them to arrive tomorrow.

SUPPOSED They _____ tomorrow.

8. Jane was thinking of looking for a new job, but she changed her mind.

GOING Jane _____a new job, but she changed her mind.

9. I went swimming every week when I was at secondary school and I really loved it!

TO I _____every week when I was at secondary school and I really loved it!

10. You should have submitted the essay yesterday.

SUPPOSED You _____ the essay yesterday.

V. Partial immediate post-test 5 after viewing episode 10

Complete the gap with one word only

1. John hasn't got any money. Paul hasn't got any, _____.

*Complete each sentence with **two to five words**, including the word in **bold***

2. I really want to have some ice-cream.

WISH I _____ some ice-cream.

3. It would be a good idea to take your car.

WHY _____ take your car?

4. Mum, can I go out, please?

LET _____ , please.

5. Helen prefers to become a doctor.

RATHER Helen _____ a doctor.

6. It is raining heavily, I should have brought an umbrella.

WISH I _____ an umbrella.

7. I want to go out tonight, not tomorrow.

RATHER I _____ out tonight, not tomorrow.

8. Can she borrow your dress?

LET Will you _____ your dress?

9. Neither Carol nor Anne do to the gym.

DOESN'T Carol doesn't go to the gym. Anne_____.

10. I think you should ask her to come to the movies.

WHY _____ ask her to come to the movies?

Appendix E. Post-viewing questionnaire

1. Have you watched any films and/or TV series in <u>English</u> outside of the classroom (e.g. at home, at the cinema) in the last 7 days?

☐ Yes ☐ No

2. What films and/or TV series have you watched in English in the last 7 days?

_____

3. Have you watched films and/or TV series with subtitles in the last 7 days?
If yes, specify the language of subtitles.

☐ With Catalan / Spanish subtitles ☐ Without subtitles ☐ With English subtitles

☐ With subtitles in other language: _____

4. Indicate how often you have watched films and/or TV series in English in the last 7 days.

☐ Never ☐ Once ☐ Between 2 – 3 times ☐ Between 4 – 6 times ☐ Every day

5. How much time a day approximately have you spent watching films and/or TV series in English in the last 7 days?

☐ Less than 20 minutes ☐ Between 20 – 30 minutes ☐ Between 30 minutes – 1 hour

☐ Between 1 – 3 hours ☐ More than 3 hours

6. Have you done any of the following in the last 7 days in order to improve your English while watching films and/or TV series in English?

☐ Replay scenes

☐ Stop the video to write down new vocabulary and expressions

☐ Look up unknown vocabulary and expressions in the dictionary

☐ Try to use the vocabulary and expressions from the video

☐ Pay attention to new words and expressions

☐ Nothing ☐ Other: _____

7. Do you feel that you have learnt new vocabulary / expressions / grammar / pronunciation from *The Good Place* TV series?

☐ Yes, vocabulary ☐ Yes, expressions ☐ Yes, grammar ☐ Yes, pronunciation.

☐ Yes, other_____ ☐ I do not know if I am learning

8. If yes, specify what you have learnt from *The Good Place* TV series.

_____

Appendix F. Instructions for the end of the intervention reflections

After having watched 10 episodes of the fantasy-comedy TV series *The Good Place*, now is the time to reflect on the experience.

Write an essay, with paragraphs, where you will develop the following points:

- How was the experience of watching a TV series in class?
- You watched this TV show with/without captions. Would you have preferred to watch it with subtitles (in your mother tongue) or with/without them instead? Why?

Did you expect to learn anything from watching the TV show? If so, what: grammar? vocabulary? pronunciation? fixed expressions? American/cultural knowledge.

Appendix E. Post-viewing activities

Episode 1

A. Did you see or hear these words in the episode that you have just watched?

Please choose “yes” or “no”. If you are not sure, choose “no”.

1. Rabbit	2. Toast
__Yes __No	__Yes __No
3. Afterlife	4. Inferno
__Yes __No	__Yes __No
5. Drama	6. Boop
__Yes __No	__Yes __No
7. Existence	8. Immortal
__Yes __No	__Yes __No
9. Bake	10. Crummy
__Yes __No	__Yes __No

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

1. Why does Eleanor keep saying “Fork”?

a. She is hungry

b. She can´t swear

c. She wants to be polite

2. Eleanor loves…

a. Clowns in her new house

b. Her new house

c. Seafood

3. What is NOT true about Chidi?

a. He speaks French

b. He loves clowns

c. He was raised in Senegal

4. What is Janet (informational assistant) not allowed to talk about?

a. The Bad Place

b. Personal secrets

c. The future

5. How many people make it to the Good Place?

a. Almost everyone

b. About 50% of people

c. Not many people

Episode 2

A. Did you see or hear these phrases in the episode that you have just watched?

Please choose “yes” or “no”. If you are not sure, choose “no”.

1. Get away with	2. Hit me up
__Yes __No	__Yes __No
3. Out of nowhere	4. Pull through it
__Yes __No	__Yes __No
5. Care for it	6. Keep on my feet
__Yes __No	__Yes __No
7. Pull this off	8. Cut me off
__Yes __No	__Yes __No
9. Come across us	10. Hang in there
__Yes __No	__Yes __No

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

1. What does Michael know about the chaos?

    a. Who caused it

    b. Why it happened

    c. Where it happened

2. Why did Eleanor eat pieces of paper?

    a. She was cheating

b. She was upset with her friend

c. She was drunk

3. Who is a designated driver?

a. A person who pays for everyone's taxi after a party

b. A person who doesn't drink to drive everyone home after a party

c. A person who is not allowed to drive

4. What is the main reason Chidi doesn't want to help Eleanor?

a. She is selfish

b. She swears a lot

c. She causes a lot of problems

5. Why did Eleanor clean up the rubbish at night?

a. Because she felt bad about what she had done

b. To convince Chidi to help her

c. To prove Michael she is special

Episode 3

A. Did you **see or hear** these phrases in the episode that you have just watched?

Please choose “yes” or “no”. If you are not sure, choose “no”.

1. Keep heads up	2. Put me out
__Yes __No	__Yes __No
3. Here´s the thing	4. Not to mention
__Yes __No	__Yes __No
5. Boast about it	6. Steered her away
__Yes __No	__Yes __No
7. Go along with	8. Turn it down
__Yes __No	__Yes __No
9. Drag them down	10. Twisted yourself up
__Yes __No	__Yes __No

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

1. Why does Eleanor hang out with Tahani?

a. She wants to become friends with her

b. She wants to prove Tahani is not perfect

c. It is a part of her education

2. Which of these is **not** a suggested hobby for Chidi?

a. Cartographer

b. Journalist

c. Personal assistant

3. Why does the plant die?

a. Because Tahani gave Eleanor a bad plant

b. Because it represents Eleanor’s attitude towards Tahani

c. Because it represents Eleanor’s mood

4. Why didn´t Eleanor´s boyfriend buy coffee from the nearby café?

a. Because it was not good coffee

b. Because he only liked fair trade coffee

c. Because the owner of the café was a bad person

5. Who wrote the note?

a. Tahani

b. Eleanor

c. Jianyu

Episode 4

A. Did you **see or hear** these phrases in the episode that you have just watched?

Please choose “yes” or “no”. If you are not sure, choose “no”.

1. Fill it out

__Yes __No

2. To give in

__Yes __No

3. Don’t mind me

__Yes __No

4. Came up with

__Yes __No

5. Don’t chicken out

__Yes __No

6. Hit rock bottom

__Yes __No

7. Hit my ears

__Yes __No

8. Mind your business

__Yes __No

9. Freak me out

__Yes __No

10. Eager to try

__Yes __No

B. 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 did Jason Mendoza do to earn money?

a. He was a DJ

b. He was a drug dealer

c. He was a dancer

3. What is the name of the restaurant recently opened in The Good Place?

a. The Good Plates

b. Angel Cakes

c. The Food Place

4. 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

5. 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

Episode 5

A. Did you **see or hear** these phrases in the episode that you have just watched?

Please choose “yes” or “no”. If you are not sure, choose “no”.

1. Cut it out	2. Take it away
__Yes __No	__Yes __No
3. Knock her out	4. Get away with
__Yes __No	__Yes __No
5. Put them up	6. Be up to
__Yes __No	__Yes __No
7. Turn it down	8. Do without it
__Yes __No	__Yes __No
9. Take a nap	10. Bun you out
__Yes __No	__Yes __No

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

1. Why is Tahani so upset about the rankings?

a. Because Eleanor is above her

b. Because her parents always ranked her lower than her sister

c. Because she thought she was Michael´s favorite

2. Why do Eleanor and Chidi have an argument?

a. Because he spends too much time teaching her

b. Because she never washes the dishes

c. Because she watches too much TV

3. Which of these does Tahani **not** do to try to improve her ranking?

a. Plan activities for her guests

b. Ask Michael for more points

c. Try to fix the sinkhole

4. How does Michael comfort Tahani?

a. He changes her ranking

b. He tells her she is remarkable

c. He tells her she is better than her sister

5. Whose actions fix the sinkhole?

a. Michael's

b. Tahani's

c. Eleanor's

Episode 6

A. Did you **see or hear** these phrases in the episode that you have just watched?

Please choose “yes” or “no”. If you are not sure, choose “no”.

1. Narrow it down	2. By my side
__Yes __No	__Yes __No
3. Give me a hand	4. Freaks me out
__Yes __No	__Yes __No
5. Do justice to	6. Came out in
__Yes __No	__Yes __No
7. Remember me by	8. Stand by him
__Yes __No	__Yes __No
9. Take me in	10. Do my best
__Yes __No	__Yes __No

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

1. Why does Chidi go to the spa with Tahani and Jianyu?

    a. Because he is bored without Eleanor

    b. Because Jianyu asked him to

    c. Because he wants to keep an eye on Tahani

2. Why does Eleanor try to help Michael?

a. Because she never kept her promises in the past

b. Because Chidi tells her to do it

c. Because she wants to become friends with him

3. Why does Michael mention *Friends* TV series?

a. Because it's his favourite show

b. Because he tries to learn about the concept of friendship from it

c. Because he wants to have human friends

4. Why is Tahani sad after the spa?

a. Because she thinks her relationship with Jianyu is a failure

b. Because Jianyu has bad taste in art

c. Because Chidi went to the spa with them

5. Why did Eleanor agree to babysit the dog?

a. Because she was a good friend

b. Because she loved dogs

c. Because her friend's Wi-Fi was good

Episode 7

A. Did you **see or hear** these phrases in the episode that you have just watched?

Please choose “yes” or “no”. If you are not sure, choose “no”.

1. Do it yourself	2. On my watch
__Yes __No	__Yes __No
3. Leave me out	4. Take a ride
__Yes __No	__Yes __No
5. Keep a secret	6. Shrug this off
__Yes __No	__Yes __No
7. Close it down	8. Pull this off
__Yes __No	__Yes __No
9. Living a life	10. Pay it forward
__Yes __No	__Yes __No

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

1. Why did Michael stay in the neighborhood in the first place?

a. Because he loves humans

b. Because he knew there would be problems

c. Because all the architects stay in their neighborhoods

2. What is the Eternal Shriek?

a. Michael's favourite party

b. A chilled and happy retirement

c. A life-time of torture

3. Why do Eleanor and Chidi decide to kill Janet?

a. Because they don't like Janet

b. Because she knows too much

c. Because she is the only one who can operate the train

4. Why does everybody hate moral philosophy professors?

a. Because they always tell the truth

b. Because they are boring

c. Because they wear weird boots

5. Why does Michael postpone his retirement?

a. Because there is no train

b. Because he wants to investigate the murder

c. Because he loves this neighborhood

Episode 8

A. Did you **see or hear** these phrases in the episode that you have just watched?

Please choose “yes” or “no”. If you are not sure, choose “no”.

1. Take a while __Yes __No

2. Up to me __Yes __No

3. Don’t sweat it __Yes __No

4. At first glance __Yes __No

5. See you off __Yes __No

6. Come forward to __Yes __No

7. Sorted through it __Yes __No

8. Get in on __Yes __No

9. For the record __Yes __No

10. Face the facts __Yes __No

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

1. What does the glowing cube on Michael’s desk do?

a. Measures if you are a good or bad person

b. Detects lies

c. Provides some atmosphere for the interrogation

2. Who does Michael find the most useful during the interviews?

a. Chidi

b. Tahani

c. Jianyu

3. Why did people laugh at Eleanor's flat mate?

a. Because she overreacted about her dress

b. Because she wore a ripped dress

c. Because she got angry if anyone asked to borrow her dress

4. Why does Chidi think Eleanor should stay in the Good Place?

a. Because she is a good person

b. Because she is learning

c. Because he loves her

5. Which of these does Trevor **NOT** do to Michael?

a. Give him a present

b. Hurt his feelings

c. Provide him with lunch

Episode 9

A. Did you **see or hear** these phrases in the episode that you have just watched?

Please choose “yes” or “no”. If you are not sure, choose “no”.

1. I’m all set	2. Push it through
__Yes __No	__Yes __No
3. Down the line	4. To deal with
__Yes __No	__Yes __No
5. Stand up for	6. Push it over
__Yes __No	__Yes __No
7. Back out of	8. Wrap it up
__Yes __No	__Yes __No
9. Move away from	10. Look down on
__Yes __No	__Yes __No

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

1. Why doesn’t Janet know what jalapeño poppers are?

a. Because she was rebooted

b. Because she is stupid

c. Because she knows nothing about bad things

2. Why did Eleanor like her job selling medical products?

a. Because they had good working conditions

b. Because she cared about people's health

c. Because she didn't want to be a part of a team

3. Why does Michael invite the Bad Place people to Tahani's house?

a. To prevent them from destroying the entire neighborhood

b. To put them in a good mood for negotiating

c. To sing karaoke with them

4. What of these is **NOT** true about the "real" Eleanor?

a. She grew up with a happy family

b. She fought against human trafficking

c. She loves clowns

5. How does Tahani discover Jason's secret?

a. He talks to her

b. She sees him with alcohol

c. She can hear his music

Episode 10

A. Did you **see or hear** these phrases in the episode that you have just watched? Please choose "yes" or "no". If you are not sure, choose "no".

1. Keep him awake	2. Get a do-over
__Yes __No	__Yes __No
3. Stay away from	4. Trust my gut
__Yes __No	__Yes __No
5. Look up to	6. Settle down here
__Yes __No	__Yes __No
7. Short on time	8. Get away with
__Yes __No	__Yes __No
9. Cheer myself up	10. Apart from them
__Yes __No	__Yes __No

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

1. Why do Jason and Janet get married?

a. Because they are soul mates

b. Because they are kind to each other

c. Because Janet is confused

2. Why do Tahani and Eleanor spend the day together?

a. To keep each other away from Chidi

b. To find out who their soul mates are

c. To recover their friendship

3. What are "real" Eleanor and Bambadjan doing with "fake" Eleanor?

a. Helping Eleanor realize her true feelings

b. Finding legal arguments for Eleanor

c. Hanging out as friends

4. How could Chidi have ruined his friend's wedding?

a. He forgot to get the rings

b. He couldn't choose which speech to use

c. He died one month before the wedding

5. What was the main consequence of Chidi's indecisiveness?

a. He didn't have real friends

b. He died

c. All the waiters hated him

Appendix H. Statistics Scripts

1. General Linear Models in SPSS, Chapter 5:

Model 1

```
UNIANOVA Post_test_score BY Group OPT_group WITH Working_memory
Pre_test_score

/METHOD=SSTYPE(3)

/INTERCEPT=INCLUDE

/SAVE=SRESID

/PLOT=PROFILE(OPT_group*Group Group) TYPE=LINE ERRORBAR=CI
MEANREFERENCE=NO YAXIS=AUTO

/EMMEANS=TABLES(Group) WITH(Working_memory=MEAN
Pre_test_score=MEAN)

/EMMEANS=TABLES(OPT_group) WITH(Working_memory=MEAN
Pre_test_score=MEAN) COMPARE ADJ(BONFERRONI)

/PRINT ETASQ DESCRIPTIVE PARAMETER

/CRITERIA=ALPHA(.05)

/DESIGN=Group OPT_group Working_memory Pre_test_score
Group*Working_memory.
```

Model 2

```
UNIANOVA Post_test_score BY Group LLAMA_F_groups WITH Pre_test_score

/METHOD=SSTYPE(3)

/INTERCEPT=INCLUDE

/EMMEANS=TABLES(OVERALL) WITH(Pre_test_score=MEAN)

/EMMEANS=TABLES(Group) WITH(Pre_test_score=MEAN) COMPARE
ADJ(BONFERRONI)
```

/EMMEANS=TABLES(LLAMA_F_groups) WITH(Pre_test_score=MEAN) COMPARE ADJ(BONFERRONI)

/EMMEANS=TABLES(Group*LLAMA_F_groups) WITH(Pre_test_score=MEAN) COMPARE (group) ADJ(BONFERRONI)

/PRINT ETASQ PARAMETER HOMOGENEITY

/CRITERIA=ALPHA(.05)

/DESIGN=Group Pre_test_score LLAMA_F_groups Group*LLAMA_F_group.

2. Linear Mixed Model fitted in R, Chapter 6:

model0 = lmer (Test scores/maximum possible test score ~ (1|Construction), data = dataset, REML = FALSE)

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

model2 = 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)

3. General Linear Mixed Methods Model / Repeated Measures Analysis in SPSS,

Chapter 7:

```
GENLINMIXED
  /DATA_STRUCTURE SUBJECTS=name*surname
REPEATED_MEASURES=Tipo COVARIANCE_TYPE=DIAGONAL
  /FIELDS TARGET=Feeling_learning TRIALS=NONE OFFSET=NONE
  /TARGET_OPTIONS REFERENCE=0 DISTRIBUTION=BINOMIAL
LINK=LOGIT
  /FIXED  EFFECTS=group_3 OPT_group Tipo OPT_group*Tipo
USE_INTERCEPT=TRUE
  /RANDOM USE_INTERCEPT=TRUE SUBJECTS=name
COVARIANCE_TYPE=VARIANCE_COMPONENTS
  /BUILD_OPTIONS TARGET_CATEGORY_ORDER=ASCENDING
INPUTS_CATEGORY_ORDER=ASCENDING MAX_ITERATIONS=100
    CONFIDENCE_LEVEL=95 DF_METHOD=RESIDUAL COVB=MODEL
  /EMMEANS TABLES=group_3 COMPARE=group_3 CONTRAST=PAIRWISE
   /EMMEANS TABLES=OPT_group COMPARE=OPT_group
CONTRAST=PAIRWISE
   /EMMEANS TABLES=Tipo COMPARE=Tipo CONTRAST=PAIRWISE
   /EMMEANS TABLES=OPT_group*Tipo COMPARE=OPT_group
CONTRAST=PAIRWISE
   /EMMEANS TABLES=OPT_group*Tipo COMPARE=Tipo
CONTRAST=PAIRWISE
  /EMMEANS_OPTIONS SCALE=ORIGINAL PADJUST=SEQBONFERRONI.
```

Appendix I. Students reflections about the intervention

Captions group:

ID 001: One reflection that I had when I saw them was that it was really helpful to watch TV shows in a foreign language, so it has encouraged to me to watch more. And not only in English, but also in French!

ID 002: For example, I do have more vocabulary thanks to watch The Good Place and the activities related.

ID 003: My reflections have helped me realising I have to keep working on my English, I thought that only watching TV shows in English was enough but the truth is that I need to study more, never is enough.

ID 004: I am aware my main sources of knowledge are TV series, books and social media. What I have realized is that I was not doing much effort to improve my English skills. I watch and read in English every movie, series or book, but I did not pay attention to the aspects I wanted to improve.

ID 005: In fact, one of the things that I am going to change in my learning from now on is watch more series or films in English. I consider is the action with which I have improved more skills.

ID 006: There was another exercise that made me reconsider the best way to learn English. I'm talking about watching "The Good Place". The fact that we were watching it with subtitles knowing that the other class was not is something that I found very interesting. Now, I'm not sure which is the best way to watch English TV series. In one hand, English subtitles help me recognize some new words. However, even if I listen and I can spell them, I still don't know what they mean. Hence, every time I'm

watching an English series and I don’t understand some word, I search it in the dictionary. I believe this new habit changes and improves my learning experience in a significant way.

ID 007: Also watching the television series in class has helped me watch many more series and movies in English at home (though with English subtitles). In short, I now know more dynamic learning methods that I can apply at home.

ID 008: The only thing that has changed about my habits is that I tend now to watch movies with captions, something that at first was difficult but now I’m starting to get used to, so I hope it becomes something natural to me.

ID 009: I realized that watching some movies in English was not a guarantee that I was learning and that I had to study too.

ID 010: Moreover, seeing a series in English with English subtitles has made I want to watch series only with English subtitles and at the same time taking note of the expressions or words I don't know.

ID 011: Thanks to this subject I have started to see series in English with subtitles in English, something never seen in me. So, I think that's a big change in my learning habits in English, because if I've already seen two series in English and more or less, I've found out what's going on, I can go for a third.

ID 012: Since we were learning through TV series in class, it encouraged me to change a little bit my viewing habits and it made me watch audiovisual content in English more often. Having said that, I think that once you get used to it, viewing series and movies that aren’t in its original version can sound horrible.

ID 013: I think some of the practices that we have done during the course have been useful for me to change different aspects of my day. For example, “My experience

with TV series, comedies & 'The Good Place" It has made me change the way I watch movies and series, and now I do it with English subtitles.

ID 014: For example, as I said, thanks to watching the serie The Good Place now I have a wider rangeof vocabulary, and this is exactly the reason for which I have decided to watch more series in English from now on.

ID 015: So now I think and I will see some series and films in English to learn better.

ID 016: To sum up, I liked watching the TV series, because I think it is a good learning method to watch an English series with captions

ID 017: I think it is good to watch movies and series in the original language, although for me this is complex, since I need to make several views to understand and understand, I have to see the chapters of the series first in the original language with Spanish subtitles and then with English subtitles.

ID 018: Watching "The Good Place" as one of the activities in class helped me realize that listening and reading at the same time the dialogs could help me getting used to the language and learning some new words and expressions. Nowadays, most of the TV series that I watch are subtitled and in English.

ID 019: This experience did not make me change my learning habits, since I was already aware of the benefits of captions and I have been using them for a while.

ID 020: Now thanks to the view of The Good Place, I know that watching English speaking Tv Series is a lively and enjoyable way of learning, but, that I have to take notes and search for the difficult words to take advantage of the activity.

ID 021: Watching the TV series The Good Place was a really entertaining activity and also, this experience made me confirm what I really thought before. I would

say that watching content with subtitles didn't make magic and no one could learn English just by that, as it was as simple as that. I knew that, and know I still thinking the same, if you don't put attention to it and really focus and try to learn something from it, you'll not learn anything. The learning process is not about watching something and hoping that will make you learn a foreign language, is about trying hard to learn, to know more things about the language, to search and translate the words or expressions you don't know, etc. Personally, I will still watch movies and TV shows with subtitles or captions, because I really try to learn more vocabulary, grammar and expressions, and I think that's the important part about the learning process.

ID 022: After doing the activities, I think that I have to watch more TV series and films in English with captions in Spanish and later, in English to learn more new expressions,

ID 023: Also, any of my reflections helped me to change my learning experience, I still do the same: I watch everything with captions and I write every day in English.

ID024: The humor through The Good place helped me to try to put the series and films in original version (English), with subtitles in my mother tongue.

ID 025: When it comes to watching The Good Place, although I'm pleased because I found a new series and liked it, I also believe I haven't learned anything strictly educational from it.

ID 026: The Good Place has cheer me to continue watching the series at home and adding other ones in original version with subtitles and, I began to do list with new vocabulary and expressions I saw/listen.

ID 027: "The Good Place" has helped me to give a new point of view to learn English. I didn't trust that watching series in English would help me learn the language. But now I think it is one of the fastest and funniest ways to learn a language.

ID 028: When I'm watching something, I try to figure out the meaning of the words that I don't know by the context. Then, I search them on the Internet to know if I have guessed it. So, I believe that I'm learning, at least, new vocabulary.

ID 029: I would say that this has made me learned a few new expressions (that I can remember), but definitely has also improved my listening and speaking skills. I consider that having watched The Good Place with English subtitles is the best choice, because it can result more difficult to understand everything without them, and captions in my mother tongue doesn't help me improving my English as much as it does with them in English. In addition, for me it's better to read and listen to the same at the same time.

ID 030: However, if we talk about films there would be a bit of difference due to me having watched approximately ten films in English with English or Spanish subtitles depending on what I would have found first. I would like to specify that I watched five TV series in Japanese with English subtitles, and I read one book in English. I always watch any kind of film or TV series in the original version with subtitles or captions, of course, since I don't care I just need a little text support to understand the show (in English and Japanese, other languages I need full subtitles). Anyway, I don't know if I learned anything from watching them. My logic says that is obvious that I have learned something, but I can't be sure of what exactly.

ID 031: I was pleased to watch it with captions too since that's how I always watch films and TV shows at home.

ID 032: In class we watched 10 episodes from The Good Place but it didn't change the amount of TV series that I watch every day, it remained nearly the same. We watch the series with captions in English, and I prefer it that way, because you learn more words, for example you just heard a word that you don't know the meaning or how to write it, but with English subtitles you can see it and search for its meaning.

ID 033: Luckily for me, our class group watch it with captions and that helped me to not get lost in the story. Although it was not the first time that I had watched any audiovisual product in English, as I normally do it approximately 2 hours once a week; I perceived that being focused on what was written in the screen made me dismiss some important visual aspects. Finally, I would like to add that I am going to start taking some notes and searching the meanings of words while I watch any movie or TV series in English in order to learn more vocabulary, English expressions and so on.

ID 034: I am glad to have seen "The Good Place" with English subtitles because is something that I do not usually do and it makes me learn and be focused. Also because when I start a TV series with captions in English, I cannot suddenly change them to my mother tongue, so now that I am in the second season, I keep them in English.

ID 035: We watched it with captions, and I would have preferred watching it without them because they make me nervous. I thought of changing classes and going with group B, who had no captions nor subtitles.

ID 036: From my point of view, watch the show with captions it has made it easier to understand, most especially the jokes which often were games of words and difficult references that I wouldn't have understood. Before doing this activity I was

totally convinced that watching TV in English was the most useful thing you could do to learn this language, but I have realised that is not true. If you want to be able to talk and write it, you also need to do other activities. I feel like after watching the TV show I know more, but when I want say exactly what I have learn I can't, so the learning process it hasn't been totally positive. That doesn't mean that I will stop watching TV series in English because it's something I enjoy and it's not negative.

ID 037: Regarding having learnt anything from that time, I cannot confirm it. I reckon that the learning is unconscious; you learn certain words, phrases or expressions without noticing. We watched the TV show with captions and it was excellent for me. I try to watch everything with captions, so I liked having watched The Good Place in that way.

ID 038: We have spent seven weeks watching The Good Place TV series in class, but because of my lack of habit I haven't watched much more TV shows in English at home. However, I have found as a very good experience to watch TV series in class, not only to encourage this habit from now on, but also to learn more than what I initially expected. something that has helped me to learn all these new concepts have been the English subtitles, the best option for me to watch TV shows in original version.

ID 039: I guess I have learned something. If I tell you the truth, I'm not really sure about it because I never know what example to give of what I could have learned. But, in the other hand, after years and years of English TV shows I've noticed that I write ad express myself better than I did.

ID 040: If I have to watched any English TV show with subtitles I always prefer English subtitles. Otherwise I don't believe that I learn anything. I mean, if subtitles were in catalan, I will spend all time reading the subtitles and not listening at what

characters would say. I can't do two things at the same time: reading in catalan, and listening people talking in English to finally, traduce what they said to catalan and compare it with the subtitles! I have proved it before and it is a mess. In second place, I would liked to watch it without subtitles just to try it. I've always watched them with subtitles, therefore I would have liked to discover how it is watching it without subtitles and if I would understand everything they say.

ID 041: Moreover, the fact of watching it with captions in English has been a perfect decision. I believe that the fact that the subtitles and the audio were in the same language forced us not to change the language in our mind and forced ourselves to think in English and, consequently, to improve our English.

ID 042: Watching *The Good Place* was a way of learning (especially vocabulary and expressions) without really noticing it. It became what I looked forward to every day, I got really into the show and I had to finish the season. When I watched it by myself, it was the same way as we did in class: with captions in English. This has always been my preferred way, because with subtitles in Spanish I feel like you don't learn nearly anything, since you're just reading in your mother tongue and probably not paying as much attention to what you're listening. There's also the option of watching without any captions at all. I reckon I don't have that level yet, because sometimes some accents can be a bit difficult to understand if the character speaks too fast, so I believe captions in English is the best option out of all of them.

ID 043: I watch many series and films on my daily life, but not many in English. However, I usually use it when they're not available yet in Spanish or haven't been dubbed.

That's why when I see some subtitled content in English, I use subtitles as a tool to follow the argument and recognize the words I'm not able to understand. It’s true that I can learn some grammatical constructions or some vocabulary, but I usually don’t notice these aspects, or at least I don’t remember.

ID 044: I think I've improved because the first few days I looked a lot the subtitles to understand the plot, but after the days and watching more episodes, I could focus more on their voices and the image. Watching the series with English subtitles has been the best, because I think, in one way or another, they help to improve our level of this language.

ID 045: I don't think I've spent enough hours watching series and movies in English, but even so, lately I've seen more TV shows in English than at the beginning of the university year. Doing this has helped me to better understand the English dialogues and learn new expressions and words that I didn't know before. I think it's something I should do more often, as it's a good way to improve my level of English, which is not very good.

ID 046: The experience of watching a series in class has been great. I laughed and learned a lot at the same time, it's a very entertaining way to learn a new language. Besides, the class becomes more dynamic and it's something to be thankful for when you've been listening to teachers for many hours in previous classes.

ID 047: The fact that we have seen the series "The Good Place" with English subtitles has helped me to better understand the dialogues. Words and expressions that I didn't understand well just by listening to them, I saw them written so I could understand the full meaning of the sentences. In addition, the subtitles make you learn how to write these words and expressions and enrich your vocabulary much more.

ID 048: I have to admit that I don't watch the TV in English nor in other languages. However, watching TV series in class it's a good idea because it's more entertaining than doing theoretical classes all time, considering that captions help to follow the story and don't get lost, although if it would be easier for me to watch it in my mother tongue, even if I lose the real voices of actors. But, in this case, I would have learnt anything about pronunciation. For other factors such as vocabulary and grammar, I have some difficulties to remember, at the end of an episode, the new words.

ID 049: After five weeks of watching TV series in English, I realize that I understand conversations in chapters more easily, and although, I thought I had been watching series in English longer, the change has been remarkable. The main reason why I started a series at home was because of the motivation I received when I saw *"The Good Place"* in class, which I thought was a great idea that helps us mainly to improve listening and, together with English subtitles, learn at the same time some expressions, grammar… But some moments watching *"The Good Place"* in English made me not understand some jokes, but most managed to understand them thanks to the subtitles or the context. I think that is due to my lack of habit, since I don't usually see many things in English.

ID 050: In five of the last seven weeks I watched some TV series or some movies. As I reported, I didn't spend too much time on it, just two or three hours per week. That is because I only see series or movies in English when I can't see them in Spanish. For this reason, I don't think I've learned too much because I just see them for fun and I'm not paying attention on the vocabulary or expressions.

ID 051: But in the case of The good place it was different. We saw it in class with subtitles in English. I think this was the best option for me because if I had seen it

without subtitles I couldn’t have understood so many words. In fact, I didn’t expect to learn too much from seeing a series in class. I thought it would be as when I see them in my house. But just for being in class I tried to catch some expressions or vocabulary. For example, I learned the expression “that’s so weird” which I started to use when I speak English.

ID 052: Although I watched the show with captions I would have preferred to watch it without them, so in that way I could be able to see if I could comprehend what the characters were saying and to focus more on the action that was going on, because sometimes I tend to get distracted reading all that there is on the captions and I miss a piece of the action.

ID 053: My experience of watching "The Good Place" in the class has helped me get in the habit of watching audiovisual contend each week at home. Also the captions have helped me to improve my listening level and now I have recently acquired Netflix and I notice that I start to see episodes or films without captions without problems.

ID 054: I think that the use of the captions has been an aid to resume my listening of the English, I prefer to see the movies without captions because they used to distract me. I would only use them if the level of the film is too high for me.

ID 055: The experience of watching a TV serie in class was interesting for the fact that maybe if I was alone at home I never would have seen it or if I would do it I would have been watched it with Spanish subtitles since I understand English better.

ID 056: I believe watching audiovisuals in English I especially learn pronunciation and the colloquial expressions they say. My opinion is that if I would have been writing all the vocabulary I don't know or maybe some expressions I would learn more.

ID 057: After seeing the TV series “The Good Place”, I learned that I can watch a series with subtitles in English and find out what is happening in it. It has helped me to start a new TV series at home in English with subtitles in English for the first time in my life.

ID 058: If I’m sincere, at the first, when I found out that we would see the TV series in English with subtitles in English, I didn’t like it at all, because I thought I wouldn’t understand anything, and I wanted to see it with subtitles in my mother tongue. But at time that I was watching the tv series, I realized that more or less I knew what was happening. Finally, I was grateful that it was with English subtitles.

ID 059: Although I would like to watch TV series or films in English with captions, I usually watch them with subtitles. The subtitles make me be more comfortable than captions, and this is because watching content with captions requires much more concentration. However, having watched The Good Place with captions has made that I get more used to them. Captions have helped me to understand better what I am listening, as well as make me be more connected with what characters are saying.

ID 060: It was thanks to laughs and deep philosophical thoughts that new words and expressions in English stick in my mind. Because we always learn better when we enjoy the way.

ID 061: The experience of watching *The Good Place* was great. I think that watching it with captions made me realize that they can help me with focusing on words that I’m not used to listen, even though I’d still prefer to watch other series without subtitles because I think they distract me from what’s really happening in the episode.

ID062: We saw the series in original version and with captions, which personally is how I like to watch them when I do it for pleasure, as long as they are in English, if not I prefer to see them in original version with subtitles in Spanish or Catalan

ID 063: From my point of view, what you improve is your level of understanding both for the vocabulary and the speed of understanding what you hear and read. I think having seen The Good Place in class is a very interesting exercise to start consuming audiovisual content in the original version.

ID 064: Thanks to watching the TV show in English in class I feel much more motivated to watch TV in English in my free time. But I remember the first time watching the TV show I felt nervous and embarrassed as English is hard to understand for me and it was difficult to follow the conversations. As I watched more episodes, it became easier to understand everything.

ID 065: However, I would have preferred to watch this TV show with subtitles in my mother tongue because I probably would have understood the dialogues better. Also, I could have followed better the episodes.

ID 066: During these weeks I have watched a *The Good Place* in class and the experience was so valuable because it's the first time I have seen a TV show with captions and not with subtitles or doubbed to my mother tongue. I have to say that watching a series with captions is better because you see the vocabulary written as they speak and this is good to learn a language watching the show.

ID 067: My experience watching this TV series in class made me learn more pronunciation of the American English and some things of the culture that they have and was reflected in the TV show. The captions help me to learn more about pronunciation.

ID 068: I liked the experience of watching The Good Place in class and I think that watching it with captions was a good idea because it's easier for me to follow the dialogues, especially if the characters speak too fast.

ID 069: It takes some time to get used to it but once you familiarize with it, you will not be able to watch any film or TV Series without Original Version (meaning English).

Textually enhanced captions group:

ID 070: Thanks to the last days that we have been watching a TV series called The Good Place, I have learned some new fixed expressions and new vocabulary that it's what I like to learn.

ID 071: I learnt because of the captions in other color that we focused without us realizing.

ID 072: As I said before we did not study English in the conventional way, and I learned to study with these ways. One example is the viewing of The Good Place with captions, I almost never watch any audiovisual object with captions because I thought that I would not understand, just in the case that the audiovisual object did not have subtitles in my mother tongue, and that was only a few times and in a videos of ten minutes as most. So, as I saw that I can understand most of The Good Place, I started watching TV shows with captions.

ID 073: As far as the essays reflections are concerned, viewing habits and learning through TV series, in this case through The Good Place, helped me to take some action to change my learning experience and habits. For instance, from now on I will

write down the new vocabulary, sentences structures, idioms or phrasal verb that appear in the TV series that I watch.

ID 074: Watching The Good Place encouraged me to see more series and movies on the original version. Because with the results of the last pronunciation activities I saw that my pronunciation improved. So, I believe that If I continue watching series and movies in English I will reach an American accent.

ID 075: The viewing habits and learning through TV series helped me to see that I had to change my viewing style because I used to see series in English but with Spanish subtitles but now, after this exercises I decided to see the series with English captions and I think it will help me with improving my English.

ID 076: Secondly, some reflections, like for example the "viewing habits and learning through TV series" made me think and wonder why I don't use to watch TV series in V.O. to improve my listening and pronunciation. So from that reflexion I started rewatching some of my favourite series in V.O. (with subtitles) and catching some new jokes that I didn’t realized till then, because with the translation we lose some jokes that were different in their original language.

ID 077: There is just one learning experience or habit that I have changed and it is watching TV series and humor through The Good Place. I have realized that now, when I watch TV series in English, I pay more attention on the fixed expressions and the jokes they say.

ID 078: Talking about my habits or how I learned English I do not think I have made any big changes, but what I have realised is that I am not learning a lot of grammar or vocabulary from only watching TV series or movies in English with captions.

ID 079: Without a doubt, watching television series entirely in English has helped me a lot to take measures to change my learning habits. Since I have seen the whole first season of "The Good Place" with captions, I will never use subtitles in my mother tongue again.

ID 080: After seeing The Good Place in class I realized that it is better to watch series and movies in their original version because dubbing is usually worse and sometimes they change the dialogues. I have continued watching audiovisual material with captions but most of the time with Spanish subtitles.

ID 081: Since I started watching The Good Place and my reflection on Viewing habits and learning through TV series (hyperlink) I have realized that watching series in original version helps to learn new vocabulary because I pay more attention at the words they say and I discover a lot of vocabulary. For that reason, now I'm starting to watch more series in original version because I didn't watch too many before.

ID 082: I loved the series and I’m still watching it on original version, which is rare for me because I used to watch everything dubbed.

ID 083: Thanks to The Good Place, my listening skills has improved a lot. Moreover, since then, I always watch TV series in English. So, during this Christmas days, I have watched four seasons of The Good Place, the second season of Lost In Space, the first season of Daybreak and two seasons of Riverdale. So, I really can say that I understand more English people than I used to.

ID 084: Moreover, I think what I have improved the most has been listening. The sessions to watch chapters of The Good Place were effective to improving and learning new vocabulary and developing the comprehension in English. For that reason, this

activity changed my habits using English subtitles in The Good Place and consuming humor through this brilliant TV series.

ID 085: By learning through the TV series I've discovered that it goes further that I thought and now I try to see as much content as I can in original versions (that is normally in English).

ID 086: Watching The Good Place helped me to change my habits as for watching series in the language that has been filmed. However, I don't watch it with captions, but with subtitles in Spanish.

ID 087: In the other hand, I have to confess that I almost never could see the sentences's results of the series The Good Place. I knew that I hardly guessed any phrase. I really understood the plot of each chapter, but rarely I learned phrasal verbs for myself. I always have needed somebody to teach me in these aspects. It isn't effective for me the indirect teaching in my level. In general, the little proves after watching The Good Place were a big disaster. Really, I was ashamed of those results.

ID 088: When it comes to my English skills I think I have imporved in the listening skills due to the TV series that we watch at class that was also helpful to improve our reading skills. That made me think that I should watch TV shows and movies with English captions instead of Spanish subtitles.

ID 089: This was by far my favorite activity and learned way more than I think I could with a TV series.

No captions group:

ID 090: Also, I have improved the vocabulary thanks to the TV Series we have been watching during the semester The Good Place, because since then, when I watch a Serio or film, I make more effort focusing on the expressions and the vocabulary I don't understand as well as the understanding of a native English speaker.

ID 091: Watching The Good Place have changed my learning habits. My group was the one analyzing the TV series without subtitles and that gave me confident to not use subtitles anymore in audiovisual content. My listening is getting used to and I feel very confident about it. Arellano Sara

ID 092: I can notice a high improvement in my language vocabulary. Maybe because of those exercises at the end of class or having seen the series "The good place" in class have made me improve quite a lot

ID 093: As I said to see the series "The good place" helped me a lot to learn in a fun way of the English language.

ID 094: Unconsciously, while listening to it, he grasped vocabulary that then sought its meaning, learned grammatical structures and learned expressions and phrases.

ID 095: Although it seems a lie to me it seemed a very good idea to learn in an easy and fun way and take it practically to my daily life and now usually I usually see series only in English if I can and with subtitles in Spanish.

ID 096: After finishing the task of viewing habits and learning through TV series, I realized that watching TV series with subtitles in English is a fun way to learn English, so from now on, I will watch TV series with English subtitles instead of doing it in Spanish.

ID 097: Now I find it easier to watch a TV series or a film without subtitles and captions, even if they are not English/American people and they talk with different accents. Maybe I easily forget the words or the grammatic used by the characters, but I understand what they say, and I can see clearly the plot. I guess I have a "better ear" now.

ID 098: I attempted to watch movies without subtitles. Since we did it in class with The Good Place, I realized I could do it at home because I did not really need the Spanish subtitles anymore.

IS 099: On the one hand, my top 3 best activities are, again, the portfolio because as I already said I think is a very good way to present our homework in an enjoyable form, watch the TV series because it made me to start watching series and films in original version.

ID 100: In terms of language skills, I have improved in the listening and speaking parts. My ears have become much more used to listening to English through watching The Good Place without subtitles. Indeed, the task Viewing habits and learning through TV series made me realize that I had to turn subtitles off much often to keep on with this marvellous learning experience.

ID 101: In fact, I had never had the habit of watching television series in English, and the fact that we were asked if we had watched some series in English, was the excuse to download Netflix and start watching series in their original language.

ID 102: Although I wrote that I would still choose the ones with subtitles in English, I did find some advantages. At least, when watching film in Spanish cinema, I can understand most of VOSE [original version subtitled in Spanish] films by listening to the voice. I think I had the ability before but I never found because I depend on the

subtitles only after leaving them can I find where my true level is. I am grateful to “Good Place”, as well as English class.

ID 103: To be honest, I would prefer to watch The Good Place with English captions for two reasons. The first one is that I’ve got a medium level of English. I was be able to understand the dialogue and what it was happening, but I believe that with captions I would understand the story better. The second reason is that I’ve got visual memory and I think that reading the dialogue I would have internalized more vocabulary.

ID 104: Watching *The Good Place* without subtitles was also a good way to start this new habit, because is a thing that most of the people in Spain doesn´t do, although I believe that it would have been also good to watch the episodes with English subtitles. I usually watch series and movies with English subtitles because I can write down any unknown word or expression, and then look for its meaning. Also, I think a better way to retain new information is by listening and looking at the same time.

ID 105: Nowadays, I don’t watch anything with subtitles and I fell great about it. Thanks to all this, I made an effort to force my listening and now I understand everything quite well.

ID 106: Normally the TV series that I see are usually in Spanish and I only watch TV series in English when they are premiere episodes and / or are not available in Spanish.

ID 107: But in spite of everything, and that I have to put subtitles in my native language I have realized that I have learned many expressions, pronunciations and in general many aspects of the English language.

ID 108: Something similar has happened to me watching the TV series "The good place". When I saw the chapters in class I felt like I was missing a lot of things,

even though I understood the general plot and most of the subplots. But in turn, although it was difficult for me to understand everything completely, I unconsciously drew many words, expressions and pronunciations that I would later incorporate into my knowledge of the English language.

ID 109: It could be that if I had seen the series TV with subtitles, at least in English, where I could see in writing what the characters were talking about, maybe I would have followed the plot more easily without missing details. Many times the characters spoke very quickly for me and I did not quite understand them. Or for example some character to have different accent to the other characters did not quite understand it at all.

ID 110: I think that it's a very good idea to watch a TV series in class because you learn many things like vocabulary, fixed expressions or pronunciation.

ID 111: I was in the group that didn't have subtitles and, personally, I would have preferred to have subtitles since I understand it better and I learn more words, expressions and, above all, pronunciation, as I do in my house when I watch TV series.