UNIVERSITAT DE BARCELONA

L1 / L2 subtitled TV series and EFL learning: A study on vocabulary acquisition and content comprehension at different proficiency levels

Ferran Gesa Vidal

L1 / L2 SUBTITLED TV SERIES AND EFL LEARNING: A STUDY ON VOCABULARY ACQUISITION AND CONTENT COMPREHENSION AT DIFFERENT PROFICIENCY LEVELS

Tesi doctoral presentada per

Ferran Gesa Vidal

com a requeriment per a l'obtenció del títol de

Doctor en Filologia Anglesa

Programa de Doctorat: *Estudis Lingüístics, Literaris i Culturals*

Línia de recerca: *Lingüística Aplicada*

Departament de Llengües i Literatures Modernes i d'Estudis Anglesos

Directora: **Dra. Immaculada Miralpeix Pujol**

Tutora: **Dra. Carmen Muñoz Lahoz**

Universitat de Barcelona

2019

ACKNOWLEDGEMENTS

During the preparation of this dissertation, I have received help and advice from many people, to whom I would like to express my gratitude.

En primer lloc, m'agradaria donar les gràcies a la supervisora de la tesi, la Dra. Imma Miralpeix. Sense el seu ajut, absolutament res de tot això hauria estat possible. Imma, des de bon principi, m'he sentit molt estimat i valorat, i això és quelcom que no es pot agrair amb paraules, i que alhora ha resultat indispensable per tal que tot aquest treball arribés a bon port. Moltes gràcies per la teva amabilitat, la teva expertesa, la teva paciència, els teus ànims i els teus consells. Gràcies també per les moltíssimes hores que has dedicat a la tesi, traient temps d'allà on no n'hi havia, per tal que mai féssim tard enlloc. No tindria prou espai ni temps per agrair-te tot l'esforç que m'has dedicat, però, si més no, volia agrair-t'ho públicament amb aquestes poques i insuficients ratlles. De tot cor, treballar amb tu ha estat un veritable plaer!

A la Dra. Carmen Muñoz, moltes gràcies per confiar en mi i donar-me l'oportunitat de treballar durant més de quatre anys al costat de tot l'equip del Grup de Recerca en Adquisició de Llengües (GRAL). Gràcies també pels teus consells i la teva implicació en el disseny de l'estudi de la tesi; sense el teu suport i la teva experiència, de ben segur que aquesta tesi no hauria estat el mateix. Moltes gràcies per tot!

Al Dr. Roger Gilabert no li podré agrair suficientment l'oportunitat que em va donar, quan jo encara era alumne del Grau d'Estudis Anglesos, de començar a treballar a la Universitat de Barcelona. Moltes gràcies, Roger, per les moltes estones que hem compartit, gràcies per escoltar-me en els moments difícils i donar-me consell, gràcies per fer-me riure i compartir amb mi la teva visió del món. En definitiva, gràcies per cuidar-me com si fos un membre més de la teva família.

A la Dra. Maria del Mar Suárez, primer de tot, gràcies per accedir a recollir dades amb els teus alumnes de Comunicació Audiovisual. Gràcies també per llegir la tesi abans de dipositar-la (i trobar el més petit dels errors), però sobretot moltes mercès per les innumerables estones que hem passat junts. Estones plenes de rialles, d'intercanvi d'opinions, d'esmorzars i dinars, de viatges, que espero es segueixin repetint en el futur.

Moltes gràcies a tots i cadascun dels membres del GRAL de la Universitat de Barcelona, he après molt de tots vosaltres i hem compartit moltes estones que m'han ajudat a créixer tant acadèmicament com personalment. En especial, m'agradaria destacar l'ajuda prestada pel Dr. Joan Carles Mora que, com a coordinador de la línia de recerca del programa de doctorat, sempre ha estat disponible per resoldre tots i cada un dels dubtes que anaven sorgint. Així mateix, també m'agradaria agrair-li l'ajuda estadística que ens ha prestat al llarg de l'elaboració de la tesis.

També m'agradaria expressar la meva gratitud als membres del Departament de Llengües i Literatures Modernes i d'Estudis Anglesos de la Facultat de Filologia i Comunicació, molt especialment a les companyes i companys que han anat passant durant aquests anys per la sala MacDermott; sense elles i ells, l'elaboració de la tesi hauria estat molt més feixuga. També m'agradaria destacar el rol de la Mercedes i l'Olga, que sempre han estat disponibles quan necessitava ajuda i han compartit amb mi estones molt agradables.

Gràcies a l'equip directiu, al personal docent i als alumnes del Col·legi Lestonnac de Mollet del Vallès i de l'Institut Damià Campeny de Mataró. Gràcies per deixar-me compartir amb vosaltres tot un curs acadèmic i obrir-me les portes de les vostres aules per poder recollir dades amb els estudiants de 6è de Primària i 4t d'ESO, respectivament. Molt en especial, m'agradaria donar les gràcies a la Isabel i a la Sandra; sense elles, la recollida de dades hauria estat impossible. Als alumnes de l'assignatura 'Expressió Oral i Escrita en Anglès' del Grau de Comunicació Audiovisual de la Facultat d'Informació i Mitjans Audiovisuals de la UB, gràcies per participar en la intervenció pedagògica durant el trimestre que vau compartir amb la Maria del Mar.

I would also like to express my sincere gratitude to Dr Stuart Webb and Dr June Eyckmans for welcoming me in their institutions and giving me the chance to spend some months under their guidance. Thank you for offering interesting suggestions for my work, which have contributed to making the thesis richer and from which I have learned a lot. Thank you to Takumi and Aki for making my stay in Canada an enjoyable experience. I am also really thankful to Griet, Alexandra, Vanessa and Ana for making me feel at home during my stay in Ghent.

I would like to thank Kieran Donaghy, Dr Carmen Muñoz (again) and Dr Michael Rodgers for having been part of the follow-up committee and having given me valuable feedback and expert advice. I am thankful as well to the members of the examining committee for having accepted the invitation to be part of the board and for having read the thesis with careful attention.

I am indebted to Melissa Cokely, who shared her expertise with her students and helped me to choose the TV series for this experiment. Thank you as well for volunteering to record the audio files needed for the study. Finally, thank you for putting on the native accent to the thesis.

También me gustaría agradecer la ayuda predoctoral (BES-2014-068089) que me concedió el Ministerio de Economía y Competitividad, así como las dos ayudas de movilidad (EEBB-I-16-10987 y EEBB-I-18-12815) que me permitieron hacer sendas estancias de investigación en Canadá y Bélgica. Esta tesis tampoco habría sido posible sin los proyectos de investigación FFI2013-47616-P y FFI2016-80564-R, dirigidos por la Dra. Carmen Muñoz, y que fueron financiados por el Ministerio de Economía y Competitividad. També m'agradaria posar en valor el finançament rebut per part de la Generalitat de Catalunya, gràcies al reconeixement del GRAL com a Grup de Recerca Consolidat (2014 SGR 01089 i 2017 SGR 560).

Per últim, però no menys important, he d'expressar la meva gratitud a la meva família i amics, que m'han recolzat durant tot aquest procés. Gràcies als meus pares, Joan i Teresa, i al meu germà, Martí, per encoratjar-me a fer allò que de debò m'agrada i per donar-me suport durant tots aquests anys. Finalment, la meva vida no seria el mateix sense el Marc: milions de gràcies per tota la teva ajuda i per sempre estar al meu costat. A tots ells, que han patit quan jo patia i han gaudit quan jo gaudia, moltes gràcies!

ABSTRACT

Whether and how video viewing is beneficial for language learning has been a matter of debate over the last decades (Vanderplank, 2015). Multimedia learning theories (Mayer, 2009; Paivio, 1986) suggest that exposure to multimodal input promotes Foreign Language (FL) learning, although the Cognitive Load Theory (Sweller, 1994) defends that multimodality may hinder it. First and second language subtitles have been found to enhance content comprehension (Baltova, 1999) and vocabulary learning (Frumuselu, 2015), even if opposite results have also been reported (Rodgers, 2013). However, most of the studies so far have been conducted with adult university learners viewing short videos (e.g., Montero Perez, Peters, & Desmet, 2013) and there is virtually no research on younger less proficient learners, with sustained exposure to video viewing, and comparing it to other teaching and methodological approaches. Therefore, this dissertation aims at analysing the effects of viewing subtitled TV series on the acquisition of FL vocabulary and on the degree of content comprehension at different proficiency levels.

The present work aims at filling these gaps by exposing Catalan / Spanish English as a Foreign Language (EFL) learners at three proficiency levels (primary school –beginners–, high school –intermediate–, and university –upper-intermediate–) to subtitled TV series over an academic term (university) or year (primary and high school). At each level, participants (*N*=158) were randomly allocated to the Experimental (EG) or control group. On a weekly basis, all learners were taught a set of Target Words (TW) and completed two vocabulary tasks (a pre-task at the beginning of the session and a post-task at the end), but only those in the EGs additionally watched a subtitled episode of a TV series containing the TWs. In order to assess lexical gains, all learners were pre- and post-tested on their knowledge of TW forms and meanings at the beginning and the end of each of the three terms the study was divided into. To measure retention effects, a vocabulary delayed post-test was also administered to the primary and high-school groups eight months after the end of the pedagogical intervention. Learners in the EGs were also regularly tested on content comprehension after viewing each of the episodes.

Results on vocabulary learning show that viewing subtitled TV series was beneficial for participants' vocabulary development. However, significant differences between conditions were only revealed in primary and high school, but not at university. More significant differences were also found in the last term, indicating that exposure to subtitled TV series is more effective if it is sustained over time. Vocabulary learning through video viewing is also mediated by the participants' proficiency level, since intermediate learners benefitted from the pedagogical intervention to the greatest extent. Results from the vocabulary delayed post-test indicate that the benefits of being exposed to video viewing were not maintained in the long-term. Regarding comprehension, there was not a steady improvement throughout the intervention at any of the proficiency levels, and scores were both participant- and episode-dependent.

The findings are discussed in the light of what previous research on the topic has found, as well as multimedia learning theories and the role that proficiency level plays in learning from multimodal input. Results on video viewing are considered for further research in the field and potential advantages of this practice are also described. Therefore, not only does this thesis bring new evidence to research on multimodal input and FL learning, but it also offers new insights into EFL teaching.

RESUM

Els possibles efectes del visionat de vídeos per a l'aprenentatge de llengües han estat motiu de debat durant les últimes dècades (Vanderplank, 2015). Les teories d'aprenentatge multimèdia (Mayer, 2009; Paivio, 1986) suggereixen que l'exposició a input multimodal promou l'aprenentatge de llengües estrangeres, encara que la Teoria de la Càrrega Cognitiva (Sweller, 1994) defensa que la multimodalitat pot dificultar-lo. Més enllà d'això, s'ha demostrat que els subtítols en llengua materna i estrangera promouen un major grau de comprensió auditiva (Baltova, 1999) i l'aprenentatge de vocabulari (Frumuselu, 2015), tot i que també s'han trobat resultats oposats (Rodgers, 2013). D'altra banda, la majoria dels estudis existents s'han portat a terme amb aprenents universitaris adults visionant vídeos de curta durada (ex.: Montero Perez, Peters, & Desmet, 2013). En conseqüència, pràcticament no hi ha estudis amb participants més joves amb un nivell més limitat de llengua estrangera, amb una exposició prolongada en el temps i comparant la citada experiència amb altres tècniques i metodologies d'aprenentatge. D'aquí que aquesta tesi tingui com a objectiu analitzar els efectes de veure sèries de televisió subtitulades sobre l'adquisició de vocabulari i el grau de comprensió en llengua estrangera a diferents nivells de competència.

El treball investiga aprenents d'anglès bilingües català / castellà, amb tres nivells de competència diferents (educació primària –nivell inicial–, educació secundària –intermedi–, i universitat –intermedi-alt–), que veuen sèries de televisió subtitulades al llarg d'un trimestre (universitat) o un curs acadèmic complet (primària i secundària). A cada nivell, els participants (*N*=158) van ser dividits aleatòriament en dos grups: experimental (GE) i control. Cada setmana s'introduïen paraules noves a aprendre per tots els estudiants i aquests completaven dues tasques amb aquest vocabulari (una pre-tasca al principi de cada sessió i una post-tasca al final), però només els estudiants dels GEs veien, a més, un episodi subtitulat d'una sèrie de televisió on apareixien les paraules noves. Per tal de mesurar els guanys en vocabulari, tots els estudiants van completar, al principi i al final de cadascun dels trimestres de l'estudi, un pre- i un post-test on s'avaluava el seu coneixement de la forma i el significat de les paraules. Per tal de mesurar el grau de retenció del vocabulari après, els estudiants de primària i secundària van fer, a més, un post-test de vocabulari vuit mesos després de l'acabament de la intervenció pedagògica. Els estudiants dels GEs també eren examinats del grau de comprensió de cada episodi immediatament després de visionar-lo.

Els resultats de vocabulari mostren que veure sèries de televisió subtitulades fou beneficiós pel seu aprenentatge. Malgrat això, les diferències significatives entre grups només sorgiren a educació primària i secundària, però no a nivells superiors. Així mateix, es trobaren més diferències significatives entre els grups al darrer trimestre, indicant que l'exposició a sèries de televisió subtitulades té efectes més positius si és prolongada en el temps. L'adquisició de lèxic a través del visionat de vídeos també està relacionat amb el nivell de competència dels estudiants: els participants de nivell intermedi van ser els que més es van beneficiar de la intervenció pedagògica. Els resultats del segon post-test, però, indiquen que els beneficis del visionat de vídeos no es mantingueren a llarg termini. Pel que fa als resultats de comprensió, no es va trobar un increment significatiu en cap dels tres nivells de competència: les puntuacions obtingudes variaven segons els participants i els episodis.

Els resultats obtinguts es comparen amb els d'altres estudis previs i són interpretats en relació a les teories d'aprenentatge multimèdia i al paper del nivell de competència en l'aprenentatge amb input multimodal. Així mateix, les troballes de la tesi apunten a futures línies de recerca

en el camp i al potencial real d'aprenentatge amb materials multimèdia. Per tant, aquesta tesi ofereix no només noves evidències en la recerca sobre input multimodal i adquisició de llengües estrangeres, sinó també altres perspectives sobre l'ensenyament de l'anglès.

TABLE OF CONTENTS

INDEX OF TABLES

INDEX OF FIGURES

INDEX OF APPENDICES

INDEX OF ABBREVIATIONS

BAF	Barcelona Age Factor
BNC	British National Corpus
CCTV	Closed-Captioned Television
CEFR	Common European Framework of Reference for languages
CG	Control Group
CL	Cognitive Load
CLT	Cognitive Load Theory
COCA	Corpus of Contemporary American English
CTML	Cognitive Theory of Multimedia Learning
DCT	Dual Coding Theory
EFL	English as a Foreign Language
EG	Experimental Group
ER	Extensive Reading
ESL	English as a Second Language
FL	Foreign Language
GRAL	Grup de Recerca en Adquisició de Llengües Language Acquisition Research Group
LTM	Long-Term Memory
L1	First Language
L2	Second Language
MC	Multiple Choice
NS	Native Speaker
OPT	Oxford Placement Test
PoS	Part of Speech
PPVT	Peabody Picture Vocabulary Test
RM	Repeated Measures
RQ	Research Question
SLA	Second Language Acquisition
SUBTiLL	Subtitles in Language Learning
T/F	True or False
TL	Target Language
TW	Target Word
VKS	Vocabulary Knowledge Scale
VS	Vocabulary Size
WF	Word Family
WM	Working Memory
WPM	Words Per Minute

CHAPTER ONE - INTRODUCTION

This first chapter will give some information about the context in which the study is carried out, with special attention to the amount and type of audiovisual input that European and Spanish citizens are exposed to. It will also deal with the advantages and disadvantages offered by the two main language practices adopted in cinema and television (dubbing and subtitling). The reasons to use TV series as input for language learning will then be summarised and the motivations to conduct this doctoral dissertation will also be explained. Finally, the organization of the thesis will be outlined at the end of the chapter.

1.1. Contextualisation

Living in the 21st century implies living in a modern and globalised society that has easy access to paper-based resources and to the Internet. Although the former is still a very prominent way of transmitting information, Internet and online resources are becoming more central to people's lives. For instance, 83% of Catalan people had access to Internet at home in 2016 (Fundación Telefónica, 2016) and this figure increases exponentially if we also take into account the access using mobile phones and tablets. According to the same source, 67% of Spanish people between the ages of 16 and 74 use the Internet on an everyday basis. Among them, teenagers and adolescents are the age group who access the World Wide Web more often. They are therefore in contact with other cultures and with people who speak a language which may be unknown to many of them (e.g., by watching videos on YouTube, surfing the Internet, reading international blogs and newspapers, chatting with friends, having meetings through Skype, etc.). Since users may not be proficient enough in the language used in these sources, they might need to resort to different linguistic tools (e.g., automatic translators, online dictionaries, glossaries, subtitles, etc.) in order to understand the aural / written texts and videos.

From all the languages to which learners can be exposed to, English is predominant in all channels of information transmission and communication. This can be explained by the fact that, for the last decades, the Western society is undergoing an acculturation process led by the Anglo-American culture, which has spread out reaching many cultural domains. For instance, people nowadays can be exposed to the predominant culture in areas as diverse as leisure, ideological politics, fashion or language use (Verdú, 1996). Moreover, figures suggest that English is the most used language worldwide, with more than 743 million people speaking it as a Second Language (L2); a larger number than people having it as their First Language (L1) (Simons & Fennig, 2018). There are also many ways in which people can be exposed to English, as it has now become the world's lingua franca *par excellence*. Among others, one of the most popular ways is through watching (online) movies and television programmes. According to the report by Austin, Barnard, and Hutcheon (2016), in that year, Spanish people watched an average of 232 minutes of television per day and their Internet consumption was of 105 minutes a day. Taking a longitudinal perspective, since the early 2010s, the TV consumption in Spain has slightly decreased, whereas Internet has become a more popular tool of media consumption. Compared to other Western European countries, Spaniards consistently tend to watch more TV than other European citizens (232m vs. 223m, respectively), although Spanish Internet consumption is inferior to that of Western European society (105m vs. 128m, respectively) (Austin et al., 2016). These reports do not clarify whether Spanish and European citizens watch English-language television or are exposed to Internet contents in English when they consume online or offline media. Therefore, we do not have more specific data available of how much time consumers are in touch with a language that is not their L1.

In contrast, the EACEA Report on the use of subtitling (European Commission, 2011) provides useful information on dubbing and subtitling practices. Out of the 556 films released in Spain in 2010, 218 (i.e., 39.2%) were American productions in English, a higher proportion than those films financed by Spanish film producers (33.5%). Consequently, 66.5% of all films released in the country needed to be subtitled or dubbed so as to reach the general public. Spain is among the few European countries which opt for dubbing the movies broadcasted in cinemas, while subtitling is the preferred option in other 26 European countries such as Sweden or The Netherlands. Apart from Spain, only Italy, Germany, Austria, France and some regions in Belgium and Switzerland dubbed foreign movies (see Figure 1.1). On average, in the past few years, 53% of the films released in Spain were dubbed and 29% were both dubbed and subtitled. Among those productions with a higher box office (namely American), only 31% of them were subtitled. However, the demand for subtitling has increased in the major dubbing countries (Danan, 2015).

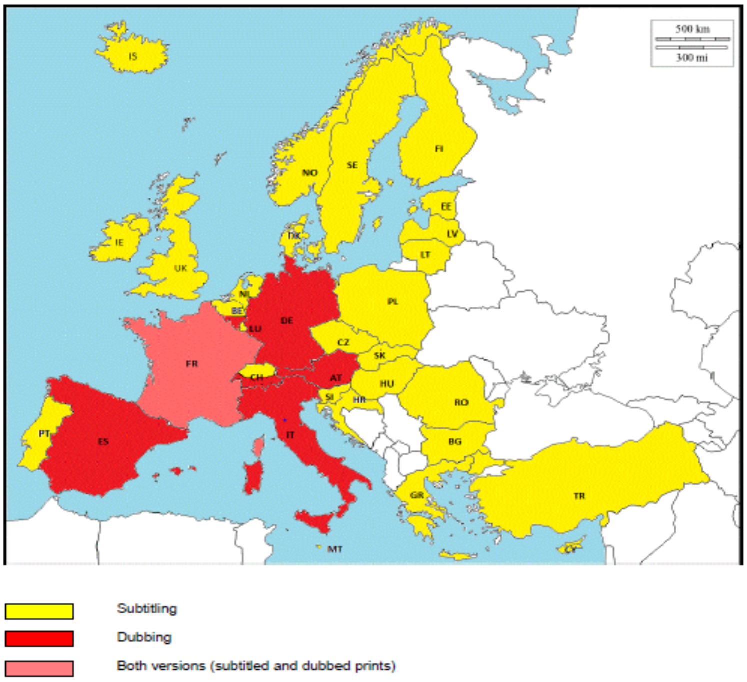


Figure 1.1 - Map of language practices in cinema (European Commission, 2011, p. 8)

A different pattern can be observed in relation to language practices in television. In this case, subtitling remains the preferred option in 16 countries, although dubbing is the dominant practice in 11 states. Besides, a new type of language practice (‘voice-over’) emerges, where an extra text is read aloud in the L1 of the audience and is put on top of the original audio, whose volume is turned down, exposing viewers to the two audios in different languages. Nonetheless, this practice is only predominant in four countries (Bulgaria, Poland, Latvia and Lithuania) and in some areas of Estonia. Finally, TV programmes aired in Luxembourg and Malta are broadcasted in their original version, most of them without being dubbed or subtitled (see Figure 1.2).

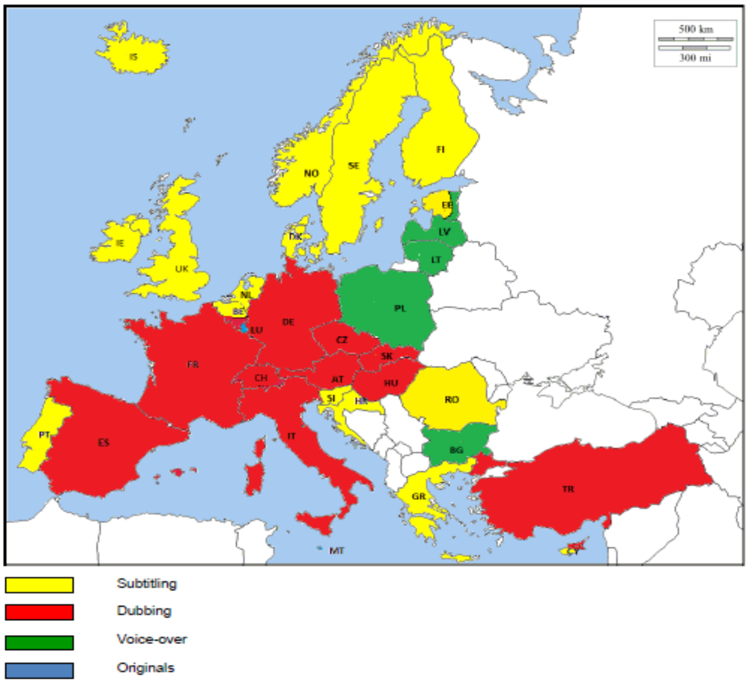


Figure 1.2 - Map of language practices in television (European Commission, 2011, p. 9)

Focusing on Spain and its television broadcasting habits, a total of 19,839 hours of TV were broadcasted in 2010, out of which only 4,046 were programmes of Spanish origin. Besides, almost 10,000 hours of the TV broadcasted in Spain were produced in the United States (50.3%); increasing the number of hours to be subtitled or dubbed up to 79.6% (European Commission, 2011).

Subtitling is usually the preferred option in small countries, with a limited budget for adapting television and cinematic contents. Larger and wealthier countries tend to opt for dubbing, as the costs are easily balanced out by the popularity of TV programmes and the income that TV broadcasters get from advertising companies and agencies. Apart from a country's size and wealth, contemporary European history and the belligerent opimate in the 1930s-1940s could also determine a country's preference for either subtitling or dubbing (Almeida & Costa, 2014). Most countries under a fascist regime (see the exception of Portugal), and with a tough censorship system, opted for a single language policy which was aimed at strengthening the feeling of national unity and the prevalence of the domestic versus the foreign. Years later, this tradition of adopting a one-language policy determined the choice of language practices in audiovisual contents (e.g., note that Germany, Italy and Spain, all three under a totalitarian regime for some years during the 20th century, now dub movies and TV programmes). Besides, the audiovisual tradition (subtitling vs. dubbing) of the country one lives in influences one's perception of the usefulness of both options: those living in 'subtitling countries' prefer subtitling whereas those living in 'dubbing countries' show a preference for the dubbed version of films and television programmes (European Commission, 2011).

Subtitling or dubbing practices have their own advantages and disadvantages. Koolstra, Peeters, and Spinhof (2002) carried out a thorough evaluation of the pros and cons of both methods analysing three different parameters: information processing, aesthetic value and learning effects (see Table 1.1 for a summary of pros and cons). Among the advantages of dubbed programmes, they argue that they are very easy to adapt, and that viewing can be combined with other tasks, as the programmes will sound familiar due to the presence of the viewers' L1. Besides, they also mention that dubbing stimulates vocabulary acquisition in one's mother tongue. Subtitling is seen as a useful tool because it condenses information into concise units, and it is perceived as being more authentic. Finally, they propose that subtitling fosters Foreign Language (FL) acquisition.

Table 1.1
Advantages and disadvantages of subtitling and dubbing (Koolstra et al., 2002, p. 344)

	Subtitling	**Dubbing**
Information processing		
Information loss through condensation	-	
Conciseness through condensation	+	
Incomplete or stretched translations		-
Redundancy and additional information	+	
Easy manipulation and censorship		-
Easy adaptation		+
Limited view on moving pictures	-	
Distracts attention from the picture	-	
Viewing easy to combine with other activities		+
Viewing possible with environmental noise	+	
Asks for high mental effort	-	
Efficiency through reading	+	
Aesthetics		
Authenticity through hearing original actors	+	
Familiarity through hearing own language		+
Overlaps with the picture	-	
Disturbs the unity of picture and image	-	
Bad translations because of lip-synchronicity		-
Bad translations because of condensation	-	
Unnaturalness through asynchronicity		-
Learning effects		
Stimulates reading development	+	
Stimulates vocabulary acquisition in own language		+
Stimulates FL acquisition	+	
Helps against developing barbarisms		+

Note. An advantage (+) of one method can be regarded as a disadvantage (-) of the other method and vice versa.

Indirectly, a country's choice between dubbing and subtitling is related to the number and level of FLs that the inhabitants master. Based on more than 5,000 answers to a questionnaire especially designed to figure out the effects of subtitling and dubbing on FL proficiency, it has been found that "the inhabitants of dubbing countries do not speak more foreign languages than those of subtitling countries in any of the three age groups [12-18, 18-25 and over 25] studied in the survey" (European Commission, 2011, p. 26). Overall, only 21% of respondents preferred watching dubbed films, whereas 30% opted for watching films without subtitles and 49% for watching films with subtitles (European Commission, 2011). Furthermore, those that speak more FLs and are more proficient in them also seem to prefer being exposed to contents subtitled in their L1 / L2 rather than watching dubbed contents.

Among all the programmes people are exposed to, TV series are becoming very popular among FL learners. In this respect, learners can be exposed to TV series in large quantities; for instance, according to a recent report, Spanish citizens watch an average of four hours of L1 television per day (Barlovento Comunicación, 2018). Similarly, most teenagers and adults in the United States spend on average five hours a day in front of the television and half of the US population is subscribed to online streaming platforms such as Netflix, HBO or Amazon Prime Video (Koblin, 2016). In the next section, the possible reasons to use TV series for language learning will be outlined.

1.2. TV series as input for language learning

In FL contexts, learners might usually look for alternative sources of input from which to acquire the Target Language (TL). This is so because students' exposure to the TL is typically limited to few hours a week and input tends to be rather scarce for effective language acquisition (Muñoz, 2008). It is known as well that, in order to achieve proficiency in a language, large amounts of comprehensible input are necessary (Krashen, 1985). Therefore, those students who want to achieve a good command of the TL may seek this input from different sources, such as in Extensive Reading (ER).

In an ER programme, learners read, either for pleasure or information (or both), large quantities of L2 texts appropriate for their proficiency level. Students are also encouraged to read at their own pace and on their own, often avoiding the use of the dictionary, and they are motivated to follow the role model of a language teacher considered to be an experienced reader (Day & Bamford, 1998). Among its various advantages, Cobb (2007) highlights the fact that ER can be beneficial for the acquisition of high-frequency words (defined by Nation, 2013 as those words belonging to the first 2,000 Word Families –WFs– and allowing for 90% lexical coverage), but this was proved to be insufficient for the learning of low-frequency words, as graded readers are mainly built with very frequent WFs. Specifically, Nation and Ming-Tzu (1999) found that it was not until level 6 (the last level in most graded readers' scales) that low-frequency words were introduced. Consequently, learners reading this kind of books may need to complement their reading with other types of input to learn less frequent words, which are the largest in number in language (Nation, 2013).

The emergence of new technologies has brought about new sources that learners can use to complement their reading. Good examples can be original songs or podcasts (which students can listen to several times until they understand the content and / or learn the lyrics); videogames (thanks to which students can learn the game's typical jargon, interact with Native Speakers –NSs– or improve their conversational skills); and limitless Internet resources, such as webpages or chats through Skype. Language learners may also watch audiovisual materials in the TL and thus be exposed to large quantities of multimodal input[1] outside school. In this respect, students of English can nowadays choose from a wide variety of audiovisual materials: videoclips on YouTube (with lyrics or not), movies in the FL (subtitled or not), and TV programmes covering different genres (documentaries, game shows, TV series, etc.).

There are different reasons why learners would resort to television programmes, especially TV series, to improve their FL skills. First, people usually have fun watching TV series, are hooked by the plot and are eager to find out how the season ends or how the relationships between characters evolve. Besides, when learners are exposed to television in the language classroom, they seem to enjoy the activity and chiefly regard television as a useful tool for studying the TL (Rodgers, 2013). This would set up the favourable conditions for language learning because, as Nation (2007) suggests, in order to learn through listening and reading (like in subtitled TV series), learners must be interested in the input they are exposed to and must be willing to understand it.

Secondly, TV series are authentic materials, defined by Gilmore (2007) as “a stretch of real language, produced by a real speaker or writer for a real audience and designed to convey a real message of some sort” (p. 98). Authentic materials have been praised by many researchers who have analysed their effects on language learning and instruction (Alipour, Gorjian, & Kouravand, 2012; Guillory, 1998; Raine, 2012; Talaván, 2007; Wang, 2012). In addition, video input and TV series are easily adaptable to all populations, as beginners can be exposed to instructional or educational videos, whereas more proficient learners can make use of the

[1] As will be explained in section 3.1.3., *multimodal input* is understood in this thesis as the simultaneous presentation of verbal, visual and auditory information, it being in the learners’ L1 or L2. As will also be shown, it has been found to promote language learning in different ways.

above-mentioned non-adapted materials, with all the advantages they have to offer (see Talaván, 2007).

Thirdly, Webb and Rodgers (2009b) found that, differently from graded readers, there was a high degree of repetition of low-frequency vocabulary in television programmes (12% of low-frequency words were repeated five or more times). This would facilitate their acquisition, since many encounters with a word are needed in order to acquire it (Pigada & Schmitt, 2006; Rott, 1999; Waring & Takaki, 2003). There is actually a good deal of repetition and linkage in television series, as the same characters tend to appear in most episodes and at least a common thread connects the same season. This allows for an increase in the knowledge that learners accumulate once they become familiar with the characters and the plot; in its turn, it also raises the possibility of learning unknown language through context cues and background knowledge (Nation, 2007). Besides, a great number of the words appearing in television series are already known to advanced language learners since the 95% lexical coverage threshold is reached at the 3k level (i.e., the most frequent 3,000 words) (Webb & Rodgers, 2009b), enabling easy understanding of a large part of the input.

In the fourth place, when exposed to video input, language learners can learn much more other than the language itself (Tschirner, 2001). Vanderplank (2010) backed up this idea by considering:

> Television or video as ideal means of showing not only authentic language but also the culture of the language being taught, both high culture (sometimes called *C* culture or cultural products) and low culture (daily customs and practices, lifestyles, sometimes referred to as little *c* culture) (p. 9).

Herron (1994) also suggests that videos allow students to "witness linguistic and cultural interactions between native speakers" (p. 190) and enable learners to be exposed to input that resembles, as much as possible, real language communication. Consequently, video input includes all the elements typical of oral language (e.g., pauses, colloquial expressions, hesitations, use of informal language, etc.), which are hard to be exposed to in a traditional language teaching context. In this way, learners can incidentally acquire certain elements that they would have missed if they had been exposed to adapted materials, where language is controlled for and specifically adapted to the level of the audience. Videos can also be used to introduce topics that can be the basis for further discussion or to raise consciousness about a problematic issue affecting society or of particular interest to students.

In the fifth place, as shown in section 1.1., L1 television consumption is probably the most popular leisure activity in the Western world. Hence, if people in Western society watch TV in their L1s on a regular basis, it is then reasonable to think they could do so in their L2 as well, bearing in mind that watching TV series in the TL is an activity that can be easily done in an in- or out-of-classroom context (Webb, 2015).

Finally, even though TV series have been part of everyday life since long ago, it was not until recently that Second Language Acquisition (SLA) researchers started focusing on their potential for language learning (e.g., Baltova, 1999; Montero Perez, Van Den Noortgate, & Desmet, 2013; Price, 1983; Winke, Gass, & Sydorenko, 2010). This growing body of research, which will be summarised in Chapters 4 and 5, has proved TV series to be a good source of language input and considers them especially beneficial for vocabulary acquisition (Rodgers, 2013) and content comprehension of authentic materials (Rodgers & Webb, 2017). For all these reasons, more research is needed on the possible effects of watching TV series on FL learning.

1.3. Why this doctoral dissertation?

There are several motivations behind this doctoral dissertation. To start with, and as pointed out in the introduction, people living in a globalised, modern society are constantly exposed to audiovisual input, sometimes in a language they are not fully familiar with. Research has shown the potential this type of input may have for FL learning (Koolstra et al., 2002), although much research is still needed. Thus, the dissertation wants to explore the possible advantages of exposure to audiovisual materials (more precisely, TV series) and analyse the extent to which they can facilitate the learning / acquisition of FLs. It was decided to focus on children, adolescents, and young adults because most studies have been conducted with adult populations at the university level and, hence, have used quite advanced language learners (e.g., Etemadi, 2012; Winke et al., 2010). Still, little is known about the effects of such input modality on other populations with a more limited proficiency in the TL.

Another motivation to conduct the study was to explore the effects of a pedagogical intervention in a classroom, which is a setting where research on multimodal input is not typically conducted. Most studies on this topic up to now have been carried out in a laboratory setting, under conditions which do not mirror the environment in which FL learning usually takes place. The present study was conducted in a classroom context with a set of pedagogical materials (not just tests) specially tailored for the participants in this context. Besides, the intervention was prolonged in time (one academic year) in order to maximise exposure and obtain longitudinal data to observe possible long-term effects. This way, participants' evolution could be traced, which is something that has already been done in this type of research, but that is, however, of the utmost importance.

In the light of what has just been outlined, the present dissertation aims at analysing vocabulary acquisition and content comprehension when viewing subtitled TV series at three proficiency levels: Grade 6 (beginners), Grade 10 (intermediate) and freshman university students (upper-intermediate). It also examines the effects of media exposure on the long-term retention of vocabulary.

1.4. Organization of the thesis

The present doctoral dissertation is divided into nine chapters. Chapter one has presented the background to the thesis and provided some context for the study, together with the motivations and the aims of the present research. Chapter two will summarise the three main theories which frame the dissertation and chapter three will focus on multimodal input for language learning. Chapter four will synthetize the research conducted specifically on vocabulary acquisition and multimodal input whereas chapter five will review the research on content comprehension of multimodal input. Next, chapter six will present the Subtitles in Language Learning (SUBTiLL) Project, the Research Questions (RQs), the participants, the design of the study, as well as the instruments and the procedure followed to collect and analyse the data. Chapter seven will thoroughly present the findings from the study, which will be interpreted and discussed in chapter eight. Finally, chapter nine will provide the reader with the final remarks, together with the limitations of the study and some ideas for future research.

CHAPTER TWO - LANGUAGE LEARNING FROM VERBAL AND NON-VERBAL INFORMATION

There are three main theories that frame the present study and help to explain the possible effects of verbal and non-verbal information on language learning: namely, the Dual Coding Theory (DCT) (Paivio, 1986, 2007), the Cognitive Theory of Multimedia Learning (CTML) (Mayer, 2009, 2014), and the Cognitive Load Theory (CLT) (Sweller, 1988, 1994). The DCT states that, in cognition, there are two independent subsystems -verbal and non-verbal- with an independent but interconnected functioning from one another. The activation of one system may lead to the stimulation of the other and this, at its turn, may trigger a greater depth of processing and better recall of information. The CTML builds upon the DCT to state that people generally learn more deeply from audiovisual explanations than from just a verbal account of information. Finally, the CLT focuses on brain's cognitive capacity and distinguishes between three types of Cognitive Load (CL), which is useful to identify ways in which learning can be enhanced or hindered. In the subsequent sections, the three theories will be described and put in relation to the aims of the present study.

2.1. Dual Coding Theory

The DCT (Paivio, 1986, 2007) is at the heart of the CTML and the CLT. According to Paivio (2007), the DCT is built upon the idea that "cognition involves the cooperative activity of two functionally independent but interconnected systems" (p. 34): one responsible for the processing of non-verbal objects and events (hereafter, imagery or non-verbal system) and the other specialized in language processing (hereafter, verbal system). The existence of the two systems mirrors how we process reality, as they can function (a) in a way that the two work at the same time but function independently from one another, (b) in a manner that allows them to work in parallel and interconnected with one another or (c) with one system activated while the other remains dormant. In (b), referential connections can link the two subsystems, as representations in one of them may activate those in the other and so initiate activity in the second one, and vice-versa. When this happens, it is thought that there is better language recall and appropriate language use (Paivio, 1986; Danan, 1992) since, according to theoreticians, "memory can be enhanced when information is stored in both codes because the effect of codes is additive" (Sadoski, 2015, p. 2). Other underlying assumptions of the DCT concern basic mental structures and mental processes. According to Clark and Paivio (1991), the above-mentioned mental structures are "associative networks of verbal and imaginal representations, and the processes concern the development and activation of those structures, including the effects of context on the spread of activation among representations" (p. 151).

Paivio's theory proposes a set of systems organized in a clear hierarchical structure. At the top general level of the structure, we can find cognitive systems that serve a symbolic or representational function; these, at their turn, are divided into verbal and non-verbal symbolic subsystems, which are both divided into sensorimotor systems (visual, auditory, haptic, taste and smell). The verbal system can receive information from the visual, auditory and haptic channels (in the form of visual words, auditory words and writing patterns), whereas the imagery system can be activated by all five sensorimotor channels. In other words, the verbal system is only stimulated when we hear and remember spoken words, see written language or even remember the touch of Braille. On the other hand, the non-verbal component of our cognitive system is activated when we are exposed to sounds coming from the environment, remember scenes or pictures or when we recall the manipulation of some types of objects. Eventually, the sensorimotor systems are filled with representation units called *logogens* or *imagens*, depending on the subsystem (verbal or non-verbal) that they belong to (Paivio, 1986, 2007).

Logogens, originally conceptualized by Morton (1969), and imagens are different in the nature of their internal structures as they originate in distinct perceptual modalities. Imagens (the basis of the imagery subsystem) correspond to (holistic parts or groups of) natural objects. Paivio (2010) defined imagens as “mental representations that give rise to conscious imagery and mediate performance in recognition, memory, language, and other functional domains” (p. 209). Imagens come from different sensory modalities so that people can be exposed to visual, auditory (representing environmental sounds), haptic (enabling us to identify felt objects) and motor imagens (guiding gestures). For instance, a visual imagen would include a room within a house and this, at its turn, within a bigger scene or landscape. On the contrary, logogens correspond to heard or spoken language, where phonemic sounds are organized into syllables and these into words and up to bigger structures, such as sentences or poems. In other words, the original definition on which Paivio based the concept of logogen was actually the abstract representations that account for word recognition when activated (Morton, 1969). Nevertheless, he adapted it to postulate that logogens are “a multimodal concept that includes auditory and visual logogens as well as input and output logogens” (Paivio, 2010, p. 210). In DCT, logogens are at the centre of language processing because they are activated in all language phenomena (recognition, memory, production and thought); Paivio even states that they can be used as a variant of “lexical representation” (Paivio, 2007, p. 37). Furthermore, logogens include everything that is learned as a chunk (it being phrases, idioms or longer units such as plays or oral stories). Overall, Paivio concludes that “DCT imagens and logogens are modality-specific internal structures that map onto the sensorimotor attributes of objects and words” (Paivio, 2007, p. 39).

Another central point in Paivio's theory is the connection between imagens and logogens. Paivio distinguishes three types of connections (or levels of meaning): representational, referential and associative (see Figure 2.1). Representational associations imply that imagens or logogens, corresponding to objects or verbal stimuli, are activated by external elements and the former are available for further processing. External elements can be stimuli that come from the outer environment or contextual information consistent with the traits of a given imagen or logogen that a person has already stored in their mind. The second type of connections (referential) are found on an upper level and are seen as the link between imagens and logogens, which enable double processing. The possible interconnection between the two systems allows for basic processes such as putting a name to pictures or creating a mental image for words one listens to. This double activation is possible thanks to the fact that, when the imagen or logogen is activated after having identified an object or a name, either one or the other activates the corresponding representation in the other subsystem (Paivio, 2007). For instance, the term *United States* may arise the image of the country's outline whereas, in the opposite direction, when somebody sees a map of the US, they could name the different states that are part of the country. Lastly, associative connections join (in the verbal channel) either same-modality words with other related words or spoken words with written words (in the case of subtitles) while, in the non-verbal channel, they either connect parts into integrated objects or these objects into scenes (without taking into account their sensory modality). Continuing with the previous example, the words *United States* may evoke other lexical items such as *North America* or cities like *New York* or *Chicago* and a map of the US could trigger another visual image of a Canadian map's outline or the memory of a particular smell or taste associated to that country. As stated before, associative connections can be intermodal and cross sensorimotor modalities; for example, as part of our everyday life, we are used to reading a text aloud while tracking the printed text (auditory and visual modalities at work) or to taking notes while listening to a lecture (auditory and haptic modalities activated) (Paivio, 1986, 2007).

In relation to the processing of verbal and non-verbal stimuli, Paivio (2007) also advocates for a multimodal dual coding brain (see Figure 2.2 with the example of *telephone*). He defends that, as objects and language come in the neural system in different forms (visual, auditory and haptic modalities), there must be "modality-specific neural representations that are activated during perception, memory, thought, and communication" (p. 142) in order to foster processing when information is received through different channels. For instance, in the example of *telephone*, information can come in through hearing (thanks to the ringtone or when somebody talks about a telephone), through sight (thanks to the actual visual representation of a phone), or through touch (thanks to touching the phone itself or the vibration some of them make). At the same time, these different representations are connected so that "telephones 'out there' can be recognized and responded to in appropriate ways" (Paivio, 2007, p. 142). This multimodal conceptualization of the brain allows for a more comprehensive model that connects the different modalities of verbal and non-verbal representations, so that any object that is felt or seen can be named and, thus, evokes images in any modality. This is especially relevant for this research project, as multimodal input can help learners to create deeper mental images: the brain is already configured to deal with multimodal information, and, according to Paivio, it is able to establish representational, referential and associative connections between information coming from different modes.

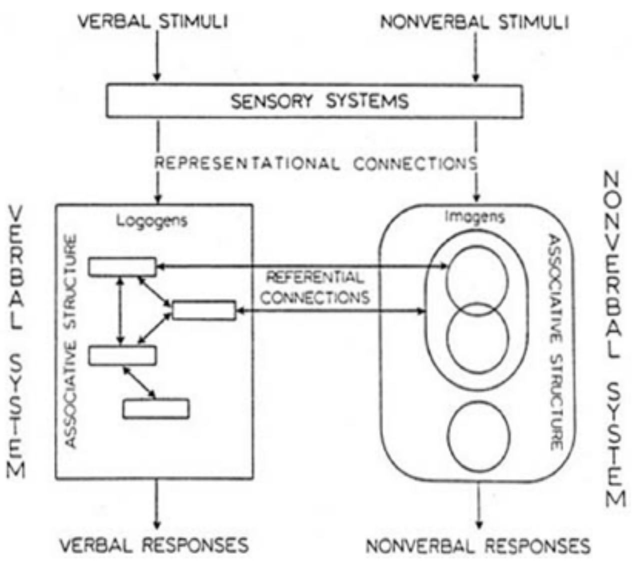


Figure 2.1 - Dual Coding Theory (Paivio, 1986, p. 67)

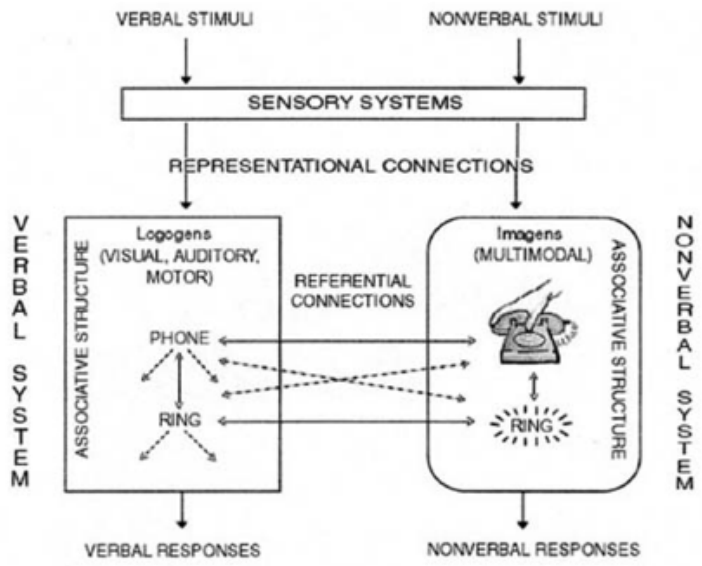


Figure 2.2 - Multimodal Dual Coding Brain (Paivio, 2007, p. 142)

Focusing on the potential of DCT for learning, Sadoski (2015) argues that any kind of information (for instance, when learning a language) is learned better if it is presented in the two modes (verbally and non-verbally). Apart from information being coded in two formats, "the connections between the two [formats] elaborate the content by giving the language concrete referents and by providing the nonverbal information with explanatory language" (p. 238). Videos provide additional visual referents for the language used and, at the same time, language can clarify the meaning of an image (for instance, seeing somebody smoking a cigar when a character in a TV series says the word *cigar* can make it easier for the learner to connect the word with its visual referent)[2]. In the light of the DCT, when people are exposed to subtitled TV series, they receive verbal information (in the form of aural words and written text) and non-verbal information (in the form of video images); this information enters the processing system through different channels (imagery and verbal subsystems) and activates both of them.

[2] Following Talaván (2012), the use of interlingual subtitles (i.e., in a language different from the one heard) adds a third type of referential connection: that between the original text and the translation provided, which is added as well to the above-mentioned referential connections already helping the language learner to process materials more efficiently.

As the two subsystems are activated, learners can build referential connections between aural, written and visual mental representations, which facilitate language learning because information is processed through the two channels. Hence, subtitles are seen as a new way to connect the verbal and the visual modes and to facilitate language comprehension and retention (Frumuselu, 2015; Talaván, 2012).

Building on the DCT's assumptions expounded, Mayer (2009) proposed the CTML, which also focuses on the idea that presenting elements in more than one mode will lead to more robust learning. Its main assumptions and principles, as well as its implications for the study, are detailed in the following section.

2.2. Cognitive Theory of Multimedia Learning

Mayer (2009) proposed the CTML based on the assumption that people learn more deeply when ideas are expressed in words and pictures rather than in words alone. However, for years, the customary way of conveying information in traditional curricular instruction has been through verbal explanations and lectures. Although this formula has a good potential for learning, Mayer suggests that an alternative form of instruction (i.e., multimedia instruction) may be more beneficial for learning purposes. He states that multimedia instruction is a very broad term that includes different activities, such as listening to an audio through speakers while viewing static images on a screen, or listening to somebody presenting a slideshow in class. In sum, multimedia instruction can be defined as the "presentation of material using both words and pictures, with the intention of promoting learning"; understanding *words* as "material presented in verbal form" and *pictures* as "material presented in a pictorial form" (it being illustrations, graphs, photos, maps, etc.) (Mayer, 2009, p. 5).

Mayer only focuses on two presentation modes (verbal and pictorial), as he suggests that learning should resemble as much as possible our cognitive functioning: assuming that there is a cognitive subsystem for words and one for pictures (as seen in DCT), if the two of them work at full capacity, there are more chances that receivers process information better. From a quantitative point of view, this would imply that receivers have double chances of learning the information in question, as they have twice as much exposure. However, this also presents some disadvantages, which have to do with our brain's limited cognitive capacity. Mayer minimises the effects of cognitive overload by saying that pictorial and verbal modes do not actually entail two different types of materials, but equivalent ways of presenting the same information. If equivalent, the two representations complement each other, and help learners to "mentally integrate corresponding pictorial and verbal representations" (Mayer, 2009, p. 7). This also implies a prevalence for either words or pictures depending on the information that needs to be transmitted, words being more adequate for harder and more difficult explanations and pictures more suitable for more intuitive materials. Finally, he concludes that it is when learners try to establish these connections between one mode and the other that they are able to reach a deeper understanding.

While the main tenet of the CTML (i.e., that people learn more deeply from a multimedia explanation than from just a verbal one) has already been discussed, the theory is also built upon three other assumptions that also need to be addressed in detail, namely, the dual-channel assumption, the limited-capacity assumption and the active-processing assumption. These three assumptions are based on the natural functioning of our brain and relate this functioning to information processing. This is why, before focusing on these assumptions, we briefly describe how the human information processing system works according to Mayer.

First, we receive input in the form of words or pictures and this input is immediately processed by two of our senses (sight and hearing). Ears can only process spoken forms of words; in contrast, eyes can process images and words (in the form of written text or subtitles). Visual and oral input is stored in sensory memory for a very brief period of time before entering Working Memory (WM), where most part of the processing takes place. When visual or oral stimuli are carefully selected for processing, they enter the WM in the form of sound images of words (for the stimulus processed through the ears) or visual images of pictures (for the stimulus processed through the eyes). Both are then organized into coherent structures to form verbal (for sound images) and pictorial (for visual images) models. At this point, both models can be integrated in the final stage within WM. This integration is also enhanced by the knowledge we have stored in our Long-Term Memory (LTM), which can be added to this newly-created knowledge. For instance, the above-mentioned integration allows learners to build up a visual representation of a rabbit when they listen to the word *rabbit*. At the same time, it also allows them to create the sound of the word *rabbit* when exposed to the image of this animal. This linkage between the two systems is what allows for a greater and deeper processing of information. If successful, this integration is moved to and stored in our LTM for its later use (see Figure 2.3).

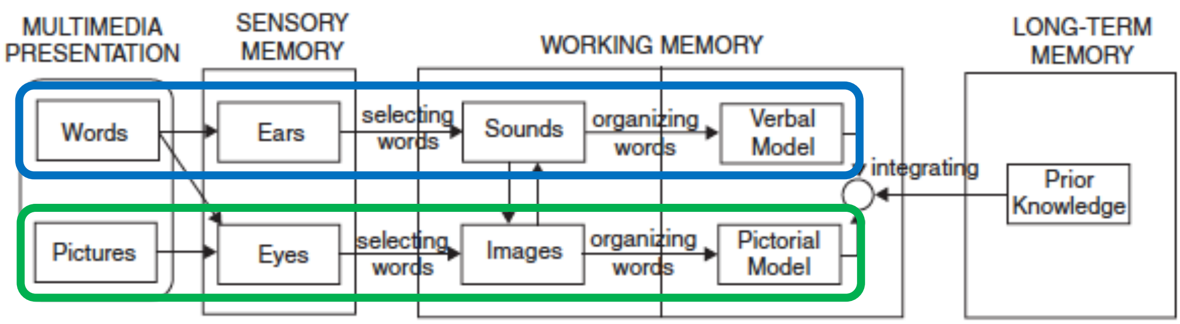


Figure 2.3 - Cognitive Theory of Multimedia Learning (Mayer, 2009, p. 61)

As can be seen in the diagram, this explanation of multimedia processing combines both systems (auditory and visual). The auditory system can be found on the top level of the diagram (highlighted in the blue box) while the visual system is situated on the lower level (highlighted in the green box). If this processing system ends up being successful, it is thanks to the three principles mentioned at the beginning of the section, which are further discussed below.

2.2.1. Assumptions of the Cognitive Theory of Multimedia Learning.

2.2.1.1. Dual-channel assumption.

Contrary to what researchers had believed for years, psychological research (e.g., Baddeley, 1986, 1992, 2000) has proved that humans possess two different channels for processing information: one for visual materials and the other for auditory materials (see Figure 2.3 above). In general terms, what is presented in the form of images or pictorial stimuli is processed through the visual channel, and what is presented in the form of words or verbal stimuli is processed through the auditory channel. However, it is possible that information which originally entered the system through one channel be later on represented in the other channel. For instance, the case of subtitles is one in which this channel change is feasible: "on-screen text may initially be processed through the visual channel because it is presented to the eyes, but an experienced reader may be able to mentally convert images into sounds, which are processed through the auditory channel" (Mayer, 2014, p. 48). As can be seen, this is closely linked to Paivio's DCT and its interconnectivity between the two cognitive subsystems.

2.2.1.2. Limited-capacity assumption.

Our brain is known to have a limited cognitive capacity (Miller, 1956) and it is only able to store information in WM for a brief period of time. Furthermore, the amount of information that can be processed in each of the channels (visual and auditory) at the same time is also limited. Thus, only few images can be simultaneously organized in a coherent visual model and the same holds true for words and coherent verbal models. As a consequence, it seems that our cognitive system can be easily overloaded if too much information is presented at the same time. Hence, learners need to carefully select the amount and type of input they receive so as to process the essential parts and form coherent auditory and visual models. It is when selecting this information that the active-processing assumption comes into play.

2.2.1.3. Active-processing assumption.

The active-processing assumption is probably the most important of the three on which the CTML is based. This assumption defends that human beings are actively engaged in processing the information they receive so as to create coherent representative models. As active processers, humans (1) select the information they want to store in memory, (2) organize this big amount of data into coherent structures and, finally, (3) integrate the newly-arrived information with prior knowledge. These three actions allow human beings to store information that can be used later in life as former knowledge when new information is processed. When focusing on each action, Mayer (2009) defines selection as “bringing material from the outside into the working-memory component of the cognitive system” (p. 70). There are two different ways in which this can be done: by selecting relevant words (i.e., transforming carefully selected spoken verbal messages into word sound bases) or by selecting relevant images (i.e., transforming a cautiously selected pictorial portion of a multimedia message into a visual image base).

The second action, organization of data, can be defined as the connection between elements (by processing them, establishing comparisons, enumerating the main elements, classifying them…) in order to form coherent verbal or pictorial models (i.e., “structured representations of selected words or images in the learner’s working memory” (Mayer, 2009, pp. 73-74). This organization aims at making sense of all the information and at trying to establish cause-and-effect chains. The organization of verbal and pictorial materials requires a previous selection of items since our capacity to process information is again limited by our cognitive resources.

Finally, integrating information could be understood as “building connections between corresponding portions of the pictorial and verbal models as well as knowledge from long-term memory” (Mayer, 2009, p. 75), the output being an integrated model that connects the two representations. Prior knowledge is seen here as an entity that helps to coordinate the integration process by making sense of the underlying structure of pictorial and verbal representations.

2.2.2. Relevant principles of the Cognitive Theory of Multimedia Learning.

Once having described the main tenet and assumptions of the CTML, it is worth reminding that the CTML is also built upon a set of theoretical principles that are said to facilitate the learning of content through multimedia instruction (Mayer, 2009, 2014):

1) Multimedia principle: deeper processing takes place when the presentation of information includes both words and pictures rather than words alone.
2) Coherence principle: aimed at eliminating extraneous material that could hinder learning.
3) Signalling principle: designed to highlight essential material.
4) Redundancy principle: based on the idea that printed text should not be added to spoken text.
5) Spatial contiguity principle: better learning is achieved when text is placed near the corresponding graphic.
6) Temporal contiguity principle: simultaneous presentation of materials leads to better learning, as opposed to successive presentations.

Apart from these six main principles, the CTML is at its turn based on four other advanced principles that research has shown to facilitate learning:

1) Segmenting principle: multimedia messages should be presented broken-down in user-paced segments rather than as a whole unit.
2) Pre-training principle: previously knowing the names and characteristics of the materials presented is beneficial for learning.
3) Modality principle: humans learn more deeply when pictures are combined with spoken rather than printed words.
4) Personalization, voice and embodiment principles: multimedia presentations are more beneficial if they reflect real humanlike conversation.

We have already discussed on the multimedia principle but, from the list above, four principles are especially relevant for the present dissertation and will be considered more in depth. They are, namely, the coherence, redundancy, pre-training and temporal contiguity principles. According to the *coherence principle*, learning is hardly ever fostered when irrelevant material is included in the presentation. As our WM has a limited operating capacity (Baddeley, 1992), it is very important to limit ourselves to include the essential material that needs to be processed in order to grasp the main idea of the multimedia presentation. Should extraneous material (i.e., not necessarily needed for understanding) be included, this would divert our attention and processing resources to this unnecessary information and would prevent us from processing the essentials, resulting in a lack of organization of the relevant material and a lack of integration between verbal and pictorial input. As Mayer suggests, the implications this principle has for teaching are clear: multimedia presentations should be short and simple, with no irrelevant sounds or images that distract learners' attention from the core of the material. Only when the essential idea has already been processed, will we include subsequent multimedia presentations to make materials clearer and to foster an in-depth learning of content and language. In relation to multimodal input (and more specifically, subtitled TV series). This implies that those contents not essentially relevant for the understanding of the plot may be unnecessary or may hinder learning.

The coherence principle is closely related to the *redundancy principle*, which claims that redundant information causes extraneous processing and limits our learning capacity. Thus, if competing sources of information (conveying the same meaning) are presented to the learner, they will not know on which to concentrate and this would result in a lack of learning. However, for redundancy to be present, the information offered through one channel needs to be very similar to some other information transmitted through another channel (for instance, by presenting exactly the same words that people read aloud printed in text).

According to Mayer, several studies (Jamet & Bohec, 2007; Mayer, Heiser, & Lonn, 2001; Moreno & Mayer, 2002; Mousavi, Low, & Sweller, 1995) have shown that those learners who are exposed to non-redundant information tend to outperform those who have been exposed to repetitive information. This may enter in contradiction with what the multimedia principle states, but Mayer (2009, p. 132) shows how to resolve this apparent paradox:

> The redundancy principle seems to suggest that two modalities are worse than one, whereas the multimedia principle seems to suggest that two modalities are better than one. The apparent discrepancy can be resolved by applying the Cognitive Theory of Multimedia Learning. The redundancy principle is based on a situation in which presenting words in two sense modalities –as print and speech– is worse than presenting words solely in one modality –as speech. A distinguishing feature in this situation is that the added on-screen text serves to overload the visual channel, which must also process the incoming animation. By contrast, the multimedia principle is based on the idea that learning can be improved when a narration is supplemented with corresponding animation. In this case, load on the visual channel is not increased because words are presented in the auditory channel.

Albeit redundancy, presenting information in alternative formats at the same time might also facilitate learning since different people learn in different ways (i.e., learners have different learning styles that need to be considered). Research (e.g., Dörnyei, 2005; Tight, 2010) has distinguished between three types of learners: visual learners, who prefer receiving information through sight and like seeing images; auditory learners, who tend to prefer information transmission through hearing; and kinaesthetic or tactile learners, who find it helpful to do activities that involve movement (e.g., underlining or walking around trying to remember content). Theoretically speaking, and in connection with the CTML, by having different presentation formats, teachers can accommodate all learning styles and foster learning. If information was just transmitted through hearing (as it is usually the case in formal classroom instruction), only auditory learners would benefit from it. Nevertheless, if information is presented through two different channels (e.g., auditory and visual), it can go through even if one of the paths is blocked due to learners' poor processing capacity. Hence, adding a third path of information transmission (i.e., the case of subtitled videos) increases the chance that this information gets to the learner if one or both paths are blocked and helps to activate prior knowledge when new information is integrated to the existing database (Talaván, 2012).

Another relevant principle for this study (as will be seen in Chapter 6) is the *pre-training principle*, by which learners who have been introduced the main concepts of the content information will perform better than those for whom the content is completely new (Kalyuga, 2005). We can then extrapolate that, if learners are not introduced to the main topic or to the characters in the multimedia materials, more of their cognitive resources will have to be allocated to actions such as learning the names and characteristics of the new items, leaving less free cognitive resources to process other information. However, when learners already have this information available, they can devote all their cognitive resources to make sense of the material to-be-learned. Thus, it may be advisable to introduce the setting or relevant vocabulary to students beforehand so that they can "devote their full cognitive capacity to building a coherent cognitive representation during learning" (Mayer, 2009, p. 198). In relation to vocabulary learning, students need to be introduced to new words if explicit learning is to be enhanced (Nation, 2013), and it is necessary to encounter a word many times before acquiring it (Pigada & Schmitt, 2006; Rott, 1999; Waring & Takaki, 2003).

Finally, the *temporal contiguity principle*, according to which “people learn more deeply from a multimedia message when corresponding animation and narration are presented simultaneously rather than successively” (Mayer & Fiorella, 2014, p. 280), is also met when viewing TV series, as viewers are presented with the images when the narration develops. Hence, TV series facilitate processing by presenting the two information types simultaneously.

As can be seen from all these principles, multimedia learning seems to apply for all kind of students and can be a powerful way to enhance language learning. Following the principles above, if multimedia materials are to be incorporated into the classroom context, caution must be taken to avoid increasing extraneous CL that may lead to poor processing. As there is a close link between CTML and the CLT, this theory is introduced below and put in relation to the theory just presented.

2.3. Cognitive Load Theory

Sweller (1988, 1994) proposed the CLT based on Baddeley’s model of WM, which distinguishes between a central executive system and two subsystems: a visuo-spatial sketchpad (to process two- and three-dimensional objects) and a phonological loop (to process verbal material) (Baddeley, 1992). As human WM has a limited capacity (e.g., it can hold between five and nine objects at a time; Miller, 1956), the CLT states that, for learning to be effective, brain’s cognitive capacity should not be overloaded. The theory, which is also related to the basis of the CTML, is specifically concerned with the cognitive resources that are used during learning and problem solving (Chandler & Sweller, 1991) and with the fact that ineffective presentation formats generate CL which impedes skill acquisition due to WM overloading. Thus, information should be presented without heavy CL so as not to hinder learning.

According to Sweller (1994), learning is guided by two basic processes: *schema acquisition* and *automation of intellectual operations*, which will ultimately reduce CL. Schema acquisition is how information is organized according to the manner it will be processed: when new information enters the processing system, it is organized around a set of schemas that facilitate its learning and processing. For example, the first time that somebody sees a dog, they will need to understand that it is a furry animal which barks and has four legs and a tail, and then organize all this information into a new schema. However, when the same person sees another dog, although a bit different from the first one as there are not two identical dogs, they will incorporate the new information into the already existing schema and thus processing speed will increase, and so learning effectiveness improve. This process will be repeated many times until the *dog* schema is fully constructed and acquired by the novice learner and it will be the basis for any future intellectual skill. As Sweller puts it, "intellectual skill based on knowledge is heavily dependent on schema acquisition" (Sweller, 1994, p. 297).

Learning also comprises automation of intellectual operations. When schemas are successfully acquired, learners need to become familiar with the new information and incorporate it into their processing system. This process, which may take time at the beginning, is called automation. Sweller further distinguishes between two different ways of processing information: controlled and automatic. Controlled processing takes place when the learner pays conscious attention to the information at hand. For instance, studying for an exam is a controlled process since learners need to be really focused on the activity if they want to successfully understand the main concepts and the relationship between them. In contrast, automatic processing does not require special attention by the learner; when something is already stored in our brain and it is done many times, it requires less cognitive demands and it can be done unconsciously, allowing attention resources to be allocated to another activity. At first, most intellectual operations tend to be controlled but, as familiarity increases, they become more automatic and need less attention resources (e.g., reading a text is an automatic operation for most literate learners although it used to be a controlled one when a child learns to read). In other words:

> This process of automation is the second major learning mechanism after schema acquisition and affects everything learned, including schemas themselves […]. It is only then [when intellectual operations become automatic] that intellectual performance can attain its full potential. Without automation, performance is slow, clumsy and prone to error. (Sweller, 1994, p. 298)

Finally, these two processes help to achieve the main goal of CLT, i.e., reduce CL. This can be achieved because, when schemas are brought into WM, they can be treated as single units and thus reduce CL and expand WM capacity. Furthermore, automation has a significant effect on WM because it helps to free cognitive resources (automatic processing requires less attention and effort within WM). One of the major functions of these two processes is to "ameliorate or even by-pass this restriction [WM's restriction on the number of discrete items it can process]" (Sweller, 1994, p. 299).

In spite of labelling them with different names, Mayer (2009) and Sweller (1999) distinguish three types of CL that need to be either reduced or boosted to improve learning: *extraneous, intrinsic* and *germane*. Extraneous CL deals with the way tasks are presented and is caused by any external element that is involved in them. It is also associated with processes that are not necessary for learning and it is the only type that can be altered by different instructional formats or methods. The reason why it should be reduced as much as possible is that it does not help to achieve the instructional goal and it prevents adequate cognitive processing and learning. Intrinsic CL (also known as essential) goes intrinsically with the characteristics of the tasks learners are asked to do and is determined by the interaction between the materials *per se* and the experience of the learner. This type of CL can be neither reduced nor boosted and it is determined by the number of elements that need to be processed at the same time and the interactivity between them. Those elements with a high interactivity will have more intrinsic CL and, consequently, will be more difficult to acquire. Eventually, the last type of CL is germane. Germane CL directly relates to learning and is responsible for schema construction and automation. When learners engage in one of these two processes, they invest more effort in genuine learning and, consequently, they can solve problems that were unsolvable before (van Merriënboer & Ayres, 2005). For these reasons, it should be boosted as much as possible.

These three types of CL should be balanced taking into account the limits of cognitive capacity: if a task's intrinsic CL is relatively high, then extraneous CL must be lowered. However, if intrinsic CL is low, a greater degree of extraneous CL can be tolerated as it will not exceed the limits of cognitive processing. Finally, the sum of the two (intrinsic and extraneous) must still leave some room for germane CL, i.e., for schema acquisition and automation of intellectual operations.

Research has also shown that there are some instructional approaches that can help to reduce extraneous CL and that are relevant for the purpose of the present study. Some of the instructional strategies that have been found to be effective are:

1) Worked examples (Kalyuga, Chandler, Tuovinen, & Sweller, 2001): conventional problems should be replaced by already known examples that need careful studying, as this focuses the learner's attention on problem states and possible solution steps, and less attention is directed towards new unessential concepts.
2) Goal-free activities (Sweller, 1999): replace conventional problems with a specific goal by a problem with a non-specific goal. This should only direct attentional resources to aspects which are essential to schema acquisition.
3) Activities based on the completion effect (van Merriënboer, Schuurman, de Croock, & Paas, 2002): completion problems seem to be effective for reducing extraneous CL. Material designers ought to give learners a partial solution that needs to be completed (instead of blank problems) as this reduces the size of the problem space.
4) Activities designed according to the modality effect (Mayer & Moreno, 2003; Sweller, 1999): unimodal sources of information (e.g., written texts or still images) should be replaced by multimodal sources of information (e.g., videos) since the latter uses both the visual and auditory processors of WM (i.e., visuo-spatial sketchpad and phonological loop) and CL is distributed between the two systems, which implies having more resources ready for processing and less chances of overloading.
5) Tasks based on the redundancy effect (Sweller, 1999): replace unnecessary sources of information with one essential source, as this will also free some cognitive resources (unnecessary information will not be processed).

Some of these CLT principles can be related to multimedia learning. To start with, Kruger, Hefer-Jordaan, and Matthew (2013) state that subtitles tend to increase CL in educational settings (mainly, content learning). They argue that "since subtitles increase extraneous or ineffective CL, it results in a reduction in germane or effective CL that is responsible for the formation of schemata" (p. 63). Research has proved that this is the case on many occasions, since adding redundant on-screen captions[3] increases extraneous processing. This tends to occur when learners try to reconcile the two verbal streams (audio and printed words) to make sure that the written forms correspond to the spoken forms and when they attempt to scan the words in the captions to the image presented above (Mayer, Lee, & Peebles, 2014). Furthermore, adding captions when content is learned in one's native language seems to be detrimental to learning as learners do not need them to process the meaning of content words (Mayer, 2009). However, and in contrast with these views, other authors argue that, in language acquisition settings, subtitles are seen as a tool to decrease extraneous CL (Mayer, 2002). It is claimed that, as the essential cognitive process of learning a language is not yet automated, redundant materials (i.e., subtitles) support the process while minimizing extra CL. In other words, subtitles and videos can work as a form of sheltered instruction (i.e., teaching academic content using the L2 at the same time learners are given some aids to help them with language processing). This helps students to access word meanings, as they can map the image with the spoken and written forms of the word. In addition, and disagreeing with previous research (Kalyuga, 2011; Mayer et al., 2001; Paas, Renkl, & Sweller, 2004), by using eye-tracking methodology, Kruger and colleagues also found that there was a significant difference in the student's percentage change of pupil diameter between those participants exposed to subtitled videos and non-subtitled videos, which was taken as an indication that those groups who were not exposed to subtitles experienced a higher CL. Based on their findings, the authors propose that "same-language subtitles in an educational context where students learn through medium ESL [English as a Second Language] reduce CL, which could be ascribed to the high intrinsic CL and the support provided to these students by subtitles" (p. 65). Furthermore, Mayer et al. (2014) showed that adding subtitles to a narrated L2 video was neither advantageous nor

[3] Subtitling mode in which both the audio and the textual support are presented in the participants' L2.

detrimental for learning. They conclude that the redundancy effect, by which subtitles could cause extraneous CL and lead to cognitive overload, is only applicable to NSs, not to FL learners. These results are in line with others showing that the presence of subtitles in FL materials enhances vocabulary acquisition and reading comprehension (Chun & Plass, 1996; Gu, 2013; Jones & Plass, 2002; Plass, Chun, Mayer, & Leutner, 1998).

In this chapter, three theories (DCT, CTML and CLT) which relate verbal and non-verbal information to language learning have been expounded. These different types of information can be transmitted through different modalities of input to which learners can be exposed: namely, unimodal, bimodal and multimodal input. These modes of input are presented in detail in the next chapter, with special attention to multimodal input, which is the focus of the study. Afterwards, the advantages and disadvantages for language learning of this input modality are also discussed.

CHAPTER THREE - THE ROLE OF INPUT IN SLA: A FOCUS ON MULTIMODAL INPUT

Input has long been at the centre of language learning research and has inspired some language learning theories. For instance, in the so-called Input Hypothesis, Krashen (1982) gives account of how human beings progress from the initial stages of language acquisition to becoming proficient speakers of any TL. Krashen distinguishes between two different levels of input in language acquisition: *i,* or the language that the learner can process at their present stage, and *i+1*, which is a bit beyond the learner's level of competence, also known as *comprehensible input*. He claims that, in order to move from *i* to *i+1*, the learner needs to be able to focus on the meaning of the unknown input (i.e., *i+1*), rather than on its form. It is by receiving input that is a bit beyond learner's level of competence (*i*+1) that they will make progress in learning a language. This progress is facilitated by context or extralinguistic information and by the already-existing knowledge stored in LTM. Instead of acquiring a structure by using it, the Input Hypothesis focuses on "going for meaning" (p. 21) as the main approach that leads to the acquisition of a given structure. Inevitably, other language learning theories have prioritised output, as in the Output Hypothesis (Swain, 1993), or interaction, as in the Interaction Hypothesis (Long, 1996), without neglecting, though, the importance of input.

3.1. Modalities of input

As we have seen, being exposed to input is, therefore, crucial for mastering any language. Input that learners are exposed to can come from different sources. Apart from more traditional ones (e.g., reading or listening), alternative sources start becoming popular (e.g., reading-while-listening, watching videos, etc.). So far, research has distinguished between three main modalities of input: unimodal, bimodal and multimodal input.

3.1.1. Unimodal input.

For years, the predominant way of information transmission to the language learner has been through one channel, namely sight (i.e., reading) or hearing (i.e., listening). Examples of unimodal input include ER, which involves learners reading books at their appropriate level, or extensive listening, according to which learners ought to receive large quantities of comprehensible spoken input with a high percentage of already familiar words and they should not be 'distracted' by other activities apart from listening to the audio file. In ER or extensive listening, vocabulary gains are typically obtained from guessing from context, as no help is provided other than the text itself (especially in an aural format, where learners cannot go back to a physical text) and, in some cases, a dictionary. Guessing from context also involves strengthening partially known vocabulary, as learners rely on a wide variety of linguistic clues (e.g., immediate and wider linguistic contexts, existing knowledge gained from other parts of a given text, morphological clues, common sense, etc.) (Webb & Nation, 2017).

Research has shown that ER seems to be beneficial for vocabulary acquisition and that certain factors may help to predict how successful it can be. For example, Waring and Takaki (2003) demonstrated that words in a FL can be learned after reading a single text of 400 different headwords, and Webb and Chang (2015a) highlighted that prior vocabulary knowledge influences, to a great extent, the amount of vocabulary learned through ER and the progress that the learner makes at all subsequent stages. Similarly, Horst, Cobb and Meara (1998) emphasized the positive effects of ER for enriching partially known vocabulary, increasing speed of access to lexical items or fostering the linkage between word form and word meaning.

Despite being less popular than reading, learning through listening also poses some advantages for vocabulary acquisition: research found that extensive listening helps to learn unfamiliar words although many encounters with the new items are necessary for successful learning to occur (Elley, 1989). For instance, van Zeeland and Schmitt (2013a) had learners listen to four short excerpts containing some non-words varying in frequency of occurrence and they took three vocabulary tests: results showed that more than 15 encounters were required for learning the form and meaning of the words. Vidal (2011) compared the vocabulary acquired by listening to academic lectures or by reading specialized texts and saw that gains were obtained in both cases; reading was a more efficient source of vocabulary acquisition (especially at lower levels), although those words learned through listening were more stable after measuring retention effects with a one-month delayed post-test. The author also highlights the importance of learners' proficiency level when acquiring vocabulary through listening or reading, since those students with a higher proficiency experienced greater gains than those at a more initial stage. Finally, Chang (2012) compared extensive to intensive listening[4] in two groups of learners, and found that the extensive listening group outperformed the intensive counterpart at the end of the intervention, indicating that students' listening competence can be enhanced when regularly exposed to a set of audiobooks without formal instruction; however, this difference was not captured when standardized vocabulary level tests were administered.

As seen above, reading and listening (as forms of unimodal input) are two major sources of vocabulary acquisition and research has proven that both are beneficial for the development of learners' proficiency and Vocabulary Size (VS). As Nation (2013) puts it, "deliberate learning and learning from output help, but they are not necessary" (p. 216), highlighting the relevance

[4] Defined by Chang (2012) as listening for specific information, for the exact words of a phrase or expression, for details or to mimic a text.

of meaning-focused input[5], without which no learning takes place. Lately, other modalities of input have become more popular in language teaching settings and discussion has arisen about the extent to which they are beneficial for FL learning.

3.1.2. Bimodal input.

In contrast to unimodal input, when learners are exposed to bimodal input, they receive information from two different channels, and therefore it enters our cognitive system through the two subsystems defined by Paivio in the DCT and Mayer in the CTML (see sections 2.1. and 2.2. for a detailed explanation). Bimodal input, best defined as the simultaneous presentation of L2 text and sound, can be considered an alternative format to ER or extensive listening. One way through which FL learners are exposed to bimodal input is when reading-while-listening: at the same time learners are reading any book that suits their proficiency level, they are also listening to the audio file reciting the text or to the teacher or role-model reader reading it aloud.

The effectiveness of bimodal input, mainly operationalised as participation in reading-while-listening programmes, has been thoroughly analysed by previous research, most of which has compared this modality of input to unimodal input. The first study to do so was that by Holobow, Lambert and Sayegh (1984), which explored the use of reversed or bimodal subtitles[6] as opposed to the usage of L2 script-only. However, instead of using video materials as expected, they proposed a series of combinations of aural dialogues and written scripts. A

[5] Learning from meaning-focused input implies that learners ought to have the opportunities to learn the language through reading and listening tasks, whose focus is on the information learners are reading or listening to (Nation, 2013).

[6] *Reversed subtitles* are understood here as the presentation of L1 audio and L2 text, whereas *bimodal subtitles* combine L2 audio and L2 text.

total of 77 Grade 5 pupils and 59 Grade 6 students, all of whom English NSs from two primary schools in Canada, were selected to be part of the study. Despite their young age, these students were quite advanced in the TL (i.e., French) as they were enrolled in an immersion course in which 40% of the instruction was delivered through the TL. Over a period of ten weeks, participants were presented with 20 different French stories (10 for Grade 5 and 10 more for Grade 6) in one of the three conditions mentioned above: (i) listening to the English dialogues while reading the French transcripts; (ii) reading and listening to the dialogues in French or (iii) simply reading the French transcripts with no audio support. For each of these 10 stories, learners were asked three general comprehension questions: two questions asking for the exact form of phrasing used in the passages and one contextual meaning enquiry. When analysing the results, participants were matched for their proficiency level, measured according to receptive vocabulary and cloze tests. In addition to checking the effects on each type of comprehension questions, a global score combining the three types was also computed for each story and submitted to multivariate tests. Results showed different behaviours depending on the condition students were allocated to: both subtitling conditions proved to be significantly better than the script-only condition although no statistical difference was observed between the two bimodal input conditions. Overall, they concluded that unimodal input is less beneficial than bimodal input since the latter "generally strengthens or enhances the verbal message, suggesting that the double modal input may be processed more deeply because attention can alternate from the auditory to the visual format or be directed along parallel visual and auditory routes simultaneously" (p. 73).

The benefits of bimodal input have been found to be especially relevant in vocabulary acquisition research. For instance, Bird and Williams (2002) compared the two modalities of input (unimodal and bimodal) in two different experiments. In the first, they presented students with four lists of (un)familiar words, while varying the modality of input learners were exposed to: some students saw the lists of Target Words (TW), whereas others heard them, and another group saw and heard the items. They were all asked to decide whether the words were familiar or unfamiliar to them and if they had seen them in a previous phase of the experiment. Results showed that participants' reaction times were faster when text was provided, suggesting that visual stimuli helped learners to make their decisions. These results revealed as well that students were not distracted by audio when text was also presented. In the second experiment, researchers explored the effectiveness of different modalities of input on repetition priming of unfamiliar words. Overall, results suggested that bimodal input boosted implicit learning of nonwords if results are compared to the sound-only condition. The authors further argued that bimodal input may be more helpful when input delivered through sound or text alone is ambiguous, as it aids speech decoding and segmentation by helping listeners visualize speech stream and clearly identify word boundaries. They eventually concluded that "bimodal input can be attended to and used to bolster both the implicit and explicit aspects of vocabulary learning" (p. 529), indicating that it can be a powerful tool to help learners with the acquisition of a FL.

Bimodal input has also been found to be more effective for vocabulary acquisition than two other approaches widely used in FL teaching: extensive listening and traditional classroom instruction. Brown, Waring and Donkaewbua (2008) exposed three groups of intermediate English as a Foreign Language (EFL) learners to three graded readers, each of which was either listened to, read or read and listened to by one of the groups. In each text, the spelling of 28 target items was modified and it constituted the body of the testing set. After completing each text, learners were given a meaning translation test and a Multiple Choice (MC) test asking for the TWs' translation. Results showed that the reading-while-listening mode promoted more learning than the other two (16% of the TWs correctly learned), followed by the reading-only method (15%) and extensive listening (2%). Although one of them led to higher vocabulary gains, no statistically significant differences were found between the first two conditions. Little decay occurred over the three months during which the study took place, although the number of items learned was relatively small. Brown et al. (2008) conclude by saying that “the nature of vocabulary learning from extensive reading or listening is more complex that can be determined from this study” (p. 158) and by opening the debate for further research on the topic.

Similar results were obtained by Webb and Chang (2015b), who also compared the outcome of an extensive reading-while-listening programme to that of regular classroom instruction. The analysis was centred on 100 TWs that were quasi-randomly chosen from ten graded readers. Students completed a meaning recognition pre- and post-test so as to calculate gains, and a delayed post-test was also administered after completing the 13 weeks of the treatment. Findings revealed that both groups (reading-while-listening and regular instruction) improved from the beginning to the end of the intervention, although the EG gained significantly more TWs than the Control Group (CG) (44.06% vs. 5.19%, respectively). Both frequency and distribution of occurrence of the TWs also seemed to be important factors, although no fixed pattern was found, and correlations did not reach significance. Nevertheless, the authors conclude that “the relatively large vocabulary learning gains in this and earlier studies examining the effects of reading multiple texts provide strong support for audio-assisted extensive reading” (p. 683).

However, it is also true that learners need to be at the right proficiency level to benefit from bimodal input; otherwise, chances are that its benefits may be lessened. In this respect, Chang

(2009) analysed the comprehension of aural texts by beginner and intermediate L2 listeners. She compared the effects of two modalities of input (unimodal vs. bimodal) on students' global comprehension and word recognition. Learners either read, listened to or both listened to and read two short stories of equal length. Results revealed that students gained a 10% more when exposed to bimodal input than when they only listened to the stories. However, proficiency was found to be a major factor affecting learners' performance since those with a higher proficiency benefited more from bimodal input than lower-level students. To account for the difference in performance, Chang argues that "reading while listening requires two language skills, and unless the listeners are well equipped with both skills, the gains may be reduced" (p. 661). Nevertheless, she advocates for using bimodal input in a classroom setting, as she founds it a good tool to foster listening proficiency in the long run, rather than a tool to improve test scores in the short term. She indicates that reading-while-listening (or bimodal input) should be introduced after learners are used to this practice and have good command of listening to stories, in order to avoid developing an inefficient approach to listening.

Finally, it is also important to note that not all studies found a positive effect of bimodal input or participation in reading-while-listening programmes. For example, Granena, Muñoz, and Tragant (2015) and Tragant, Muñoz, and Spada (2016) explored their efficacy in a bilingual setting by comparing the performance of two intact classes throughout an academic year, one of which followed a reading-while-listening approach whereas the other followed teacher-led instruction. Questionnaires, comprehension tests and writing tests were administered to participants at the beginning and end of the intervention programme and similar linguistic gains were found for both groups; hence, no significant differences were found. Based on the data of the questionnaires, the authors also suggest that L2 learning under bimodal input instruction is linked to L1 reading factors (e.g., attitude towards reading or reading environment at home) and other background aspects (e.g., parents' educational level).

In sum, most of the studies on bimodal input agree that exposing learners to text and sound simultaneously seems to be beneficial for their learning, especially if comprehension and vocabulary acquisition are evaluated. Nevertheless, research is very scarce, and these results are not conclusive. Research on multimodal input is introduced in the next section, focusing especially on its possible benefits for language learning.

3.1.3. Multimodal input.

Defined as the simultaneous presentation of written input and video information, multimodal input is the focus of this doctoral dissertation. It has been a field of study in foreign and SLA since the early 1980s. At that time, Price (1983) investigated whether exposure to captioned television improved or hindered video comprehension and found that the majority of the 500 participants involved in the study benefited from captioning.

The first models connecting multimodal input and language learning date back to Vanderplank (1990), when he proposed his own model to explain the benefits of viewing captioned television. According to it, attention plays a central role in such process and is the main device needed for learning. Following the model, learners should watch TV consciously, systematically and reflectively to later on adapt (for their own purposes) some of the language they have been exposed to. Then, they should also adopt the adapted language to produce it correctly in appropriate situations. However, he makes it clear that "adopting" language does not necessarily imply that it has been “taken in” (i.e., assimilating it to one’s linguistic competence) (p. 229) (see Figure 3.1).

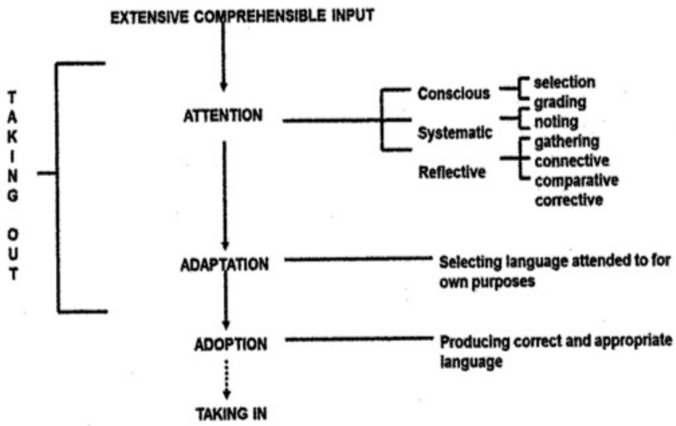


Figure 3.1 - “Paying attention to the words” model (Vanderplank, 1990, p. 229). Reproduced in Vanderplank (2016, p. 64).

Other authors have also theorized about the use of video in the language classroom, highlighting that video provides models and examples of the lexicon that speakers use to talk about a given topic. Although this purpose can also be served by still images, they do not provide the audience with the opportunity of seeing how the speaker interacts linguistically with the interlocutor, thus fostering and resembling real-life communication and favouring situated learning (Tschirner, 2001). This learning is “grounded in context at all levels from phonology to syntax and from discourse to pragmatic and sociocultural structures” (p. 318). In addition, the form-meaning link, so relevant to vocabulary learning (Nation, 2013; Webb & Nation, 2017), can be better established through the use of digital video thanks to the visual support dynamic images give to aural texts. As regards vocabulary learning, dynamic images and digital video allow audiovisual texts to be treated similarly to written texts, as learners can listen to a fragment several times (or even slow spoken language down) and more attention can be paid to unknown aspects of sign and meaning; this way, more discrete and less noticeable language features can be noticed, processed, and maybe learned (Tschirner, 2001).

Webb (2015) defends the validity of extensive viewing as a supplement to ER or extensive listening and to reading-while-listening to help to fill the need for greater L2 input. Extensive viewing involves "regular silent uninterrupted viewing of L2 television inside and outside of the classroom" (p. 159). According to the author, in-class video viewing can guide independent out-of-class viewing after having raised awareness of the benefits it has for language learning and having developed the needed strategies to support comprehension. However, he also claims that the benefits of extensive viewing may be greater in the long-term, when accumulation of video input is more considerable. Webb (2015) and Webb and Nation (2017) outline six principles that any extensive viewing programme should follow for it to be effective:

(1) The language learning benefits must be clear to anyone involved (e.g., students, parents, teachers), contradicting the popular belief that television can solely be a form of entertainment and not a source of language learning.
(2) Learners should be at the appropriate level to understand the videos chosen; if exposed to incomprehensible input, they will not experience a pleasurable viewing, and this may discourage them from further viewing. It is advisable that learners master the first 2,000 words of the TL before beginning any independent extensive viewing programme.
(3) Listening comprehension should be supported. As improved comprehension may lead to a more pleasurable viewing, it is recommended that learners watch different episodes of the same television programme, as they can gain knowledge about the setting, characters, common storyline, etc. which can aid comprehension of subsequent episodes. Other ways of fostering comprehension could be pre-teaching keywords or including captions or L2 subtitles.
(4) Precise comprehension should be a goal rather than a requirement. As comprehension gradually increases over time, learners should be told not to frustrate if they do not fully

understand the episodes. They ought to be reminded as well that the goal of extensive viewing is to reach a comprehension level that allows for enjoyable L2 television viewing.

(5) Classroom-based extensive viewing should guide out-of-class viewing. The former can serve as an example of how things should be done at home. Besides, another important aim of in-class viewing is to show to students that they can watch FL television autonomously, without the help of the teacher.

(6) Learners should watch L2 television as much as possible, as language gains will only be substantial if learners keep on watching television at home (in-class viewing will always have to be limited).

Webb (2015, p. 166) further argues:

> It is important that classroom-based extensive viewing does not get broken down into watching short segments with the learners focused on completing intensive learning activities. This will move the nature of viewing away from comprehension of meaning-focused input and toward a more language-focused orientation. Instead, pre-learning activities that aim to support comprehension and meaning focused post-viewing activities that have the objective of consolidating knowledge may be most useful.

Following principle (3), a way to support comprehension when exposed to multimodal input is including subtitles as one of the three sources of information that learners receive: dynamic images or videos, which are processed in the non-verbal subsystem according to Paivio's DCT, audio (any form of aural text in the video) and textual support (subtitles), processed in the verbal cognitive subsystem. The next section describes all possible forms of textual support that we may find in videos, together with what research has seen so far on the efficacy of this support for language learning and comprehension.

3.1.3.1. Types of subtitles.

Although sometimes used with the same meaning, as is also done in the present dissertation, subtitling and captioning are two terms to refer to slightly different forms of on-screen text appearing in videos. The main difference between subtitling and captioning depends on how much paralinguistic information is included in the transcription.

(Closed) *captioning* was originally developed for the deaf and hard-of-hearing, who experience trouble in understanding the dialogue and / or narration of television programmes. Captions are in the same language as the audio, and they show what is being said in the written form. Part of everyday life since 1990, when the US Congress passed the "Television Decoder Circuitry Act" requiring all TV sets for sale in the United States to have a caption decoder (Parks, 1994), closed captioning rapidly became mainstream. Captioning implies that not only audio content (dialogue) is transcribed but also paralinguistic information is also described for the audience (e.g., name of the speaker, manner of speaking, sound effects, songs) as it is assumed that the audience has problems with the decoding of the oral message they are being exposed to. In contrast, when the term *subtitling* is used, it refers to "rendering in writing the translation into a TL of the original dialogue exchanges uttered by the different speakers, as well as of all other verbal information that is transmitted visually (letters, banners, inserts) or aurally (lyrics, voices off)" (Díaz Cintas, 2010, p. 344). To exemplify this distinction between captioning and subtitling, one could think about someone whispering in a TV series: whereas in a closed captioning condition, the textual support would explicitly reflect that somebody is whispering between brackets (i.e., [whispering]) followed by the dialogue, in the subtitling condition only the words uttered would be translated, omitting the fact that the person is whispering, as the receiver can already notice the change in tone and pitch. Moreover, as captioning does not imply any kind of translation from the source language, the transcription is done on a more literal basis and, typically, no deletion or insertions are made when transforming aural to written text. However, when it comes to subtitling, the original text can be reduced and the language used adapted, bearing in mind that they "must strive the essence of what is said" (Díaz Cintas, 2010, p. 346) and have to be semantically and syntactically meaningful. In this respect, there are different forms of subtitling:

i) *Standard subtitling*: the audio is presented in the L2 or the FL of the audience, whereas the textual support is in the viewers' mother tongue. It helps input to become comprehensible (Danan, 2004) and it is the most widely used technique when audiovisual content is broadcasted abroad, since it helps viewers to establish the form-meaning link between the L2 word presented in an aural form and the L1 translation (Talaván, 2012). This type of subtitles has proved especially beneficial for low-level learners as it allows them to access authentic input they would not be able to follow without the L1 support (Koolstra & Beentjes, 1999).

ii) *Reversed subtitling*: the audio is in the audience's L1 and the subtitles are in the L2 or FL. Although it is less common than standard subtitling, some authors agree on the fact that it helps to acquire unknown FL vocabulary (Danan, 1992; d'Ydewalle & Pavakanun, 1997), as students benefit from listening to their own language and create connections between what they hear and what they see on the on-screen text (Talaván, 2012).

These two types of subtitles (i and ii) are also referred to with the umbrella term of *interlingual subtitles* because the language of the audio and that of the text do not coincide.

iii) *Bimodal* or *intralingual subtitling*: in this form of subtitles both the audio and the text are presented in the same language. Caimi (2006) highlights that intralingual subtitles are beneficial for language acquisition provided that they are linguistically loyal to the original dialogue and "appropriately tailored to the semantic and pragmatic markedness of the plot, speed of images and scenes" (p. 91). They are thought to have a positive impact on comprehension and vocabulary learning. Throughout this dissertation, *intralingual subtitles* and *captions* will be used to refer to the same thing, that is, the presentation of L2 or FL audio and text.

iv) Other types of textual support include *dual* or *bilingual subtitles*, in which subtitles are simultaneously presented in two different languages (L1 and L2). They have been found to be positive to develop productive vocabulary due to the fact that "when having the translation available, users are aware of the grammatically correct, well-punctuated, and unambiguous written form in both L1 and L2" (García, 2017, p. 477). This type of subtitles is very rare and only used in certain geographical areas where two or more languages are spoken (e.g., in Finland with Finnish and Swedish or in Jordan with Arabic and Hebrew).

v) *Keyword captions* are another, although less popular, type of captioning in which only the essential information is presented in a captioned format. For instance, Guillory (1998) modified the original captions to only include what NSs considered to be keywords (around 14% of the transcript). She concludes that keyword captions are an effective tool for information transmission and are also beneficial for comprehension as they "require less reading for learners, and the appearance of individual keywords in captions calls particular attention to specific content in the video […] without having to read word after word on the screen" (p. 104). Montero Perez, Peters, Clarebout, and Desmet (2014) also explored the efficacy of keyword captions and found that the resulting salience did not bring about greater vocabulary gains than in a full captioning condition. In a subsequent study (Montero Perez, Peters, & Desmet, 2018), differences between full captioning and keyword captioning were only found at the recognition level, indicating that they did not facilitate access to TWs' meanings (for more information on these studies, see Chapters 4 and 5).

vi) Finally, Kovacs and Miller (2014) developed a software that creates what they called *Smart Subtitles*, which allow the learner hover over words to show the definitions and pronunciation of those they do not know. Furthermore, learners can also get the translation of a whole fragment of the script. Although the results the authors report are very positive (participants exposed to *Smart Subtitles* correctly defined twice as many words compared to the traditional captioning group), more research is needed on this new tool. A similar study was carried out by Montero Perez et al. (2018). In this study, they exposed learners to different types of captioning, one of them being glossed captions, which allowed participants to pause the video and access the L1 translation of a set of keywords preselected by the researchers. Results showed that this newly-developed form of captioning proved to be the most beneficial out of the four conditions included in the design of the study (glossed captioning, keyword captioning, full captioning and no captioning) and participants allocated to the aforementioned condition significantly outperformed the others in all the tests administered (covering both form recognition and meaning recall). Results also indicated that glossed captions are a good tool to increase learners' lexical coverage and to access TWs' meanings. The authors conclude that "the provision of access to meaning while watching audio-visual material should be strongly encouraged because it not only promotes students' form recognition but also facilitates the construction of form-meaning connections" (p. 22).

As there are many types of subtitles, it is then reasonable to think that some will be better than others for certain areas of language learning. In this respect, research has started to explore which subtitling mode (mainly, L1 vs. L2 subtitles) leads to better comprehension and higher vocabulary acquisition at the same proficiency level (e.g., Pujadas, forthcoming). Next, the advantages and disadvantages of multimodal input for language learning will be discussed.

3.1.3.2. (Dis)advantages of multimodal input for L2 language learning.

First of all, it should be noted that different types of subtitles lead to a different processing of multimodal input. d'Ydewalle and De Bruycker (2007) analysed eye movement behaviour on two subtitling conditions: standard and reversed. Overall, they found that reversed subtitles are skipped more and fixated less than standard subtitles, maybe due to the fact that learners do not need this extra textual support as the audio is already in their L1. However, they also found that subtitles are read no matter the language they are presented in, thus evidencing automatic reading behaviour whenever subtitles are available. In a more recent study, Bisson, van Heuven, Conklin, and Tunney (2014a) thoroughly investigated processing and reading behaviour when participants are exposed to standard, reversed and intralingual subtitles in a movie. Although the main goal of the study was analysing reading behaviour, authors also looked at incidental vocabulary acquisition in each of the subtitling conditions. A total of 54 English participants with no previous knowledge of Dutch were divided into four conditions: control condition (L2 audio and no subtitles), intralingual (Dutch audio and subtitles), standard (Dutch audio and English subtitles) and reversed subtitling (English audio and Dutch subtitles).

A fifth group took the vocabulary test but did not watch the movie. Students were shown a 25-minute film excerpt and their gaze was recorded using an eye tracker; afterwards, all students regardless of the experimental condition were given a meaning recognition vocabulary test. Results showed a difference in reading behaviour between the intralingual condition and the other conditions, mainly because participants were reading in an unknown language. Furthermore, the reversed condition required the least processing, since learners did not need to resort to textual help because the audio was in English. However, automatic reading of subtitles was proven in all the experimental conditions with available textual support. Authors hypothesized that this behaviour could be explained by the saliency of subtitles or the extent to which the information presented in the subtitles needed to be processed by the participants. Finally, no between-group differences were found on the vocabulary test.

If subtitles are read and processed automatically regardless of the type, it is then reasonable to believe that their presence will bring along some advantages, but also disadvantages. Danan (2004) points out that captions are the key to understand large amounts of authentic aural input. Subtitling should also lead to better listening comprehension in face-to-face interaction with NSs or FL learners. According to Danan, subtitles bridge the gap between reading skills, which in FL learning settings tend to be more developed, and listening comprehension skills, which, in traditional classroom settings, are more likely to be neglected and less developed.

To those identified by Danan (2004), Talaván (2012) adds three more advantages of subtitles, regardless of the type. First, multisemiotic information received through subtitles adds to the visual associations taking place in memory and to the benefits of translation already inherent in multimodal input. Second, content comprehension of the message delivered is facilitated, which is essential in order to develop communicative competence. Third, viewing can be conceptualized as a more intellectual activity, which may favour audience's motivation and interest while it facilitates the understanding of humour in the L2.

Other advantages of subtitles and multimodal input for language learning have to do with the fact that learners show positive attitudes towards subtitled video viewing because the support of on-screen text results in positive reinforcement of what learners think they hear (Mariotti, 2015; Pujadas & Muñoz, 2017; Rodgers, 2013; Vanderplank, 1988; Wang, 2012; Williams & Thorne, 2000). In addition, processing of subtitles does not seem to pose important challenges for learners once they are used to them (e.g., Bisson et al., 2014a; d'Ydewalle, Praet, Verfaillie, & Van Rensbergen, 1991; Winke et al., 2010).

However, Talaván (2012) also reckons some of the limitations entailed in the use of subtitles. For example, students can miss some of the information presented aurally as their attention will be inevitably directed towards the written support. According to the author, this can be solved by preparing activities to make sure that students use the subtitles with a clear aim (e.g., noting down words or expressions uttered but not transcribed). Furthermore, as mentioned in Chapter 2, the presence of subtitles may lead to cognitive overload (specially in low-level students) and turn the above-mentioned advantages into disadvantages, which teachers need to be aware of and address.

The efficacy of subtitling and multimodal input for comprehension and learning can also be affected by learners' age and proficiency. In this sense, Vanderplank (1988, p. 280) proposes the existence of a "threshold language proficiency level", set at the intermediate level, which learners need to attain if they want to benefit from subtitling. He claims that below the intermediate level, subtitles may hinder language comprehension and acquisition because, even with this textual support, language presented in television or films may be beyond learners' grammatical and lexical level, or learners' reading skills might be too poor to be able to read the subtitles before they disappear. On the opposite end, the author claims that, for learners at an advanced level, subtitles may be distracting since they may slow down their reading and listening abilities. Danan (2004) also claims that captions may not be suitable for all ages and language proficiencies. For instance, she mentions that captions may only be beneficial for beginners if the materials have been carefully adapted to their interlanguage level. To overcome this lack of competence in the TL, Danan mentions some of the successful learning strategies that facilitate language acquisition through subtitled video viewing; for example, planning a viewing session with activities related to the clip, monitoring the watching by assessing one's own comprehension, or judging the effectiveness of the strategy employed (cf. Thompson & Rubin, 1996). Age and TL proficiency might also play a role in the processing of subtitles. As expected, processing different types of information is not an easy task for learners who do not have a sufficiently high level of the TL and may result in cognitive overload (Mayer et al., 2014). However, it is also true that d'Ydewalle and De Bruycker (2007) found no substantial differences between children's and adults' reading behaviour; even if children took longer to shift the attention towards the subtitling area and also showed longer fixations in the text.

The possible benefits of subtitling for L2 language learning have been studied in relation to different aspects, such as: vocabulary acquisition, listening comprehension, literacy development, the acquisition of grammar, foreign speech perception, speech segmentation and speech performance. Therefore, we briefly summarise in the following paragraphs the main findings in these areas (research results on vocabulary and comprehension will be expanded in Chapters 4 and 5). To start with, Vanderplank conducted two studies in 1988 and 1990 in which he analysed the effects on vocabulary uptake of watching teletext subtitles during a period from nine to twelve weeks. The first study asked 15 high-intermediate EFL learners to watch nine hours of teletext subtitled television and note down their reactions to this experience, as well as whether any words or expressions had caught their attention while watching television. Preliminary findings of the study indicated that all participants found subtitles beneficial for their lexical development and learned different strategies to cope with them. They also reported the noticing of some new words, although no controlled vocabulary tests were administered and their vocabulary progress could not be measured. In the subsequent study (Vanderplank, 1990), a similar informal approach was used: another group of 15 students watched four hours of teletext subtitled TV each week for three months and completed a series of receptive and productive tasks. The author selected a wide range of BBC programmes for this purpose. However, when teletext subtitles were not available, a written transcript of the programmes was used, which participants had to read while watching the programmes. Findings were similar to those of the previous experiment, although participants were unable to recall or produce any language used in the television programmes.

A recent state-of-the-art (Vanderplank, 2010) shows that subtitles have also been found to be beneficial for students' listening comprehension and their development of literacy. In connection with this, Neuman and Koskinen (1992) also explored whether captioned TV was a good tool for learning language and science concepts in one of the participants' L1s. They selected 129 middle-school bilingual children, most of them from Asian descent, who were subsequently divided in four different groups (captioned TV, uncaptioned TV, reading and listening to text, and textbook only) and received different types of input for nine weeks. Researchers preselected a total of 90 items that were the focus of the study and administered vocabulary and sentence anomaly tests. Data revealed that those learners exposed to science lectures through television outperformed the other two conditions (reading and listening and textbook only), although significant differences between captioned and uncaptioned conditions were not found. They concluded that "captioning presented a particularly rich language environment which enabled students to incidentally learn words through context as they developed concepts in science" (p. 27).

Regarding the acquisition of grammar, Van Lommel, Laenen, and d'Ydewalle (2006) explored the efficacy of subtitled TV programmes on learners' grammar acquisition with two experiments. In the first, participants watched a short cartoon subtitled in an unknown language (Esperanto) and were asked to complete a test which included 40 MC questions based on the grammar rules of the TL. Half of the students in the two age groups (11 and 17 years old) were explicitly told the grammatical rules the day before watching the movie, while the others were not. Results indicated that a prior presentation of the grammatical rules benefitted participants, as this group's score was significantly higher; however, "watching the movie does not in itself produce [grammatical] acquisition" (p. 249), as the gains were rather small. In the second experiment, half of the participants (different to those who volunteered for Experiment 1) were told to look for rules as applied to the movie, whereas the others were not made any grammatical reference. Both groups, though, were exposed to a video in Esperanto subtitled in L1 Dutch. Results indicated that those presented with prior grammatical information did better than those who did not receive explicit instructions. However, and similar to Experiment 1, exposure to videos in itself did not result in grammatical acquisition, which led the authors to claim that grammatical acquisition from limited video viewing may not be possible unless some explicit grammatical instruction is provided beforehand.

In relation to the influence of bimodal and multimodal input on speech perception, Mitterer and McQueen (2009) studied whether subtitles support perceptual learning of foreign speech.

A group of 121 participants watched either a 25-minute episode of an Australian sitcom or an excerpt of a Scottish movie with English (L2) subtitles, Dutch (L1) subtitles or no subtitles. After watching either the sitcom or the film, learners were asked to repeat back 160 excerpts in total, but only 80 came from the videos (40 from each): hence, half of them were new whereas the other half had already been heard. Results showed that familiarity with the accent (i.e., Australian or Scottish) helped learners to segment more words of the clip that matched the exposure they had had. Besides, the language of the subtitles (either in L1 or L2) proved to be a determinant point in their adaptation to unfamiliar accents: those who had been exposed to English subtitles repeated fragments much better than participants exposed to Dutch subtitles, who found it difficult to repeat back unfamiliar accented utterances. They concluded that L2 subtitles actually helped learners to tune in to unfamiliar accents, while L1 subtitles hindered such adaptation (this same idea is also mentioned in Vanderplank, 2015).

Drawing on this much-cited study, Charles and Trenkic (2015) looked at how good FL learners were at segmenting speech from six different genres of television programmes. Participants were asked to repeat what they had heard in 60 short excerpts (10 from each genre). Statistical analyses revealed that students missed on average a 30% of the words uttered and significant differences were found depending on the genre, films being the most challenging type of TV programme. In a second experiment, the authors aimed at exploring whether repeated viewing helped students to improve their speech segmentation ability performance. A group of 12 participants was divided into three conditions: viewing documentaries with sound and subtitles, sound only or subtitles only. Following a pre- and post-test design, learners were exposed to two 30-minute clips during two consecutive weeks (one clip each week). At the beginning and end of the experiment, they took a similar speech segmentation test as in the previous experiment. Results revealed that all participants significantly improved over time, regardless of the condition they were assigned to, although the multimodal input group showed the most pronounced improvement, proving that "watching programmes with subtitles does not seem to take attention away from listening" (p. 195). Therefore, authors concluded that this is a good practice to improve L2 speech segmentation ability.

Eventually, subtitling has also been found to have an effect on speech performance, as shown by Borrás and Lafayette (1994). Their study analysed French learners' communicative performance through multimedia tasks. Participants were pooled from four French university classes and allocated to four conditions, differing between them in terms of task complexity level (low vs. high) and the presence of subtitles (subtitled videos vs. unsubtitled videos). Participants' oral performance was assessed by measuring effectiveness, accuracy, organization and fluency, and a compound measure was calculated based on these four parameters. After viewing a video segment, learners were tested on their comprehension and they were given some time to think about the task they had to solve. Afterwards, they recorded the oral speech sample based on the notes taken. A statistically significant effect for subtitling was observed, implying that those who watched the fragment with subtitles scored higher than the others in all the four aspects analysed (and, consequently, in the compound measure). Authors concluded that, apart from being a good way to foster comprehension, "having the opportunity to see and control subtitles, as opposed to not having that opportunity, results in […] subsequent better productive use of foreign language" (p. 70).

As has been seen in this section, learners' automatic reading behaviour of captions or subtitles has been proved by different studies. Likewise, it has been shown that captions or subtitles are processed regardless of the other sources of input they may be competing with, although this may affect viewers' attention towards other visual and aural information. Textual support is also processed differently depending on its language (L1 or L2). Moreover, subtitles seem to aid literacy development, speech perception, speech segmentation and speech performance, although results so far show that subtitling per se is not very helpful in the learning of grammar. The two areas in which subtitles seem to benefit FL learning to a greater extent are vocabulary acquisition and listening comprehension, and research in these areas will be thoroughly analysed in the following chapters.

CHAPTER FOUR - VOCABULARY ACQUISITION THROUGH MULTIMODAL INPUT

In this chapter, information about the learning of vocabulary will be provided. First, in section 4.1., we will pay special attention to the distinction between incidental and intentional L2 vocabulary learning. Some word- and learner-related factors affecting L2 vocabulary learning will be briefly explained too. Next, results of different studies focusing on vocabulary acquisition through multimodal input will be presented according to the experimental design adopted. In the first group, in section 4.2., we find studies comparing vocabulary learning when students watch videos and when they do not. The second group, in section 4.3., includes those studies comparing subtitled vs. non-subtitled video viewing. Finally, the third group, in section 4.4., introduces the studies which looked at the effectiveness of different subtitling techniques (without including any uncaptioned condition). In sections 4.3. and 4.4., studies are also grouped according to the proficiency level of the participants (advanced, intermediate and beginner learners) and the effects of multimodal input observed (positive, unclear or neutral and negative effects).

4.1. How L2 vocabulary is learned

The terms *incidental* and *intentional* vocabulary learning have been widely used in the field (see for example Ellis, 1999; Huckin & Coady, 1999; Hulstijn, 2001; Malone, 2018) to distinguish between vocabulary learning processes. Traditionally, incidental vocabulary learning has been defined as that acquired "as a by-product of other cognitive exercises involving comprehension" or any other task (Gass, 1999, p. 319), whereas intentional learning is said to take place when the specific goal of the programme is to gain lexical knowledge of new and partially-known words, typically with an explicit focus on the task (Schmitt, 2008). On the other hand, others defend that the difference between incidental and intentional vocabulary learning relies on the role of learners' intention when acquiring the new lexical items (Hulstijn, 2003; Rieder, 2003). When it comes to research, though, "telling or not telling students that they will be tested afterwards on their knowledge is the critical operational feature distinguishing incidental and intentional learning" (Hulstijn, 2001, pp. 267-268). However, some question the distinction between *incidental* and *intentional* vocabulary research (Bruton, García López, & Esquiliche Mesa, 2011). They claim that any activity conducted in a classroom context or for research purposes, due to the mere presence of a task, test or the researcher, or for being different to what learners are used to doing, will somehow be intentional.

In terms of learning outcome, the vocabulary learned incidentally and intentionally is quite unbalanced, as the opportunities for learning in immersion contexts with lots of input are greater. However, teaching words deliberately (for instance, providing students with word-lists or explaining the meaning of unknown or polysemous words) is also a very effective way to quickly enlarge learners' VS. In addition, the vocabulary learning rate of the form-meaning connection in intentional contexts tends to be faster than in incidental contexts, where more repeated exposures are needed, although it remains unclear whether this is also the case for long-term learning of vocabulary (Webb & Nation, 2017). Intentional vocabulary learning is also said to lead to higher chances of reaching the productive mastery threshold and to better retention levels than incidental acquisition. However, as intentional learning takes place in classroom contexts, instruction time is limited and a precise and careful selection of the words to be taught needs to be made. Therefore, deliberate / explicit teaching of vocabulary is seen as a wondrous complement to out-of-class incidental learning (Nation, 1990, 2013; Schmitt, 2008).

Regarding incidental vocabulary learning, Huckin and Coady (1999) reviewed the bulk of research conducted on the topic and concluded that, in line with insights from several studies (Gass, 1999; Nation, 2013; Webb & Nation, 2017), incidental vocabulary learning is never purely incidental: without attention resources being allocated to the new lexical items, there will be no learning; that is why the alternative term *non-focused learning* (instead of 'incidental') may be more suitable when referring to this type of process. However, as the term *incidental* has been traditionally used in research, it will also be used in this dissertation (as opposed to 'non-focused').

In terms of requirements, a basic receptive VS of at least 3,000 WFs (with an optimal size of 5,000 WFs representing 98% coverage) is needed to be able to guess from context and incidentally learn new vocabulary (Nation, 2006, 2013). Yet there seems to be no consensus on the number of encounters needed to learn new words, depending very much on the characteristics of the TW itself (whether it is a cognate, its morphological features, etc.) and other factors such as the presence of contextual cues or the extent to which the learner is interested in learning the item. Furthermore, Huckin and Coady (1999) also claim that incidental vocabulary learning can be complemented and supported by using global and local learning strategies (e.g., graphemic identification, resorting to dictionaries or using broader contextual meanings). Textual enhancement and highlighting techniques have also been shown to have positive effects on incidental vocabulary learning, although this has the risk of not being an incidental process anymore. Research has shown too that there is a tendency to incidentally learn words when they are recognized as cognates between the learners' L1 and the TL, when there is a significant exposure to the FL, and when related L2 words belonging to the same semantic family are already known (Tonzar, Lotto, & Job, 2009). It has been shown as well that some background information (Pulido, 2000, 2003) and content familiarity (Pulido, 2004) help vocabulary learning rates to boost.

Early studies have also suggested many activities that have the potential to increase the amount of vocabulary that is learned incidentally. As said in previous chapters, one of the most popular techniques to enlarge one's vocabulary is ER (Horst et al., 1998; Nagy, Herman, & Anderson, 1985), in which students learn new words as a by-product of another task (in this case, graded book reading). However, this is not the solely method fostering incidental vocabulary acquisition research has identified. More recently, extensive listening (Chang, 2012) or extensive viewing (Webb, 2015) have also set the ground for further research on the topic.

Talking about intentional vocabulary learning, Schmitt (2008) and Webb and Nation (2017) outlined different principles that should guide in-class vocabulary teaching, such as taking advantage of meaning-focused input and meaning-focused output, and promoting learner autonomy so as to help learners to cope with the large amounts of input they are exposed to. They highlight that establishing the form-meaning link is the first and essential step in vocabulary learning and that, once the link is created, many opportunities to consolidate it are crucial, in the same way that repeated exposures to the same lexical item will be needed to gain knowledge of different aspects of the same TW (for instance, derivational affixes and collocations).

Nevertheless, it is important to emphasise the necessity of both incidental and intentional approaches. The two are not mutually exclusive; on the contrary, research has shown that merging both conduces to fruitful long-lasting learning. As Schmitt (2008, pp. 353-354) puts it:

> It is also clear that intentional and incidental approaches are not only complementary, but positively require each other. It is impossible in explicit teaching to recycle words adequately and to teach all of the contextual types of word knowledge [...], and so exposure to a great deal of reading and listening is necessary for consolidation and enhancement of explicitly taught lexical items. Conversely, words acquired by incidental learning are unlikely to be learned to a productive level, and so the additional attention that comes from intentional learning may be required to push them to this level of mastery.

It is important to clarify that, as all activities in this dissertation took place in a classroom context with explicit focus on TWs, all learning gains derived from the intervention in the present study will be catalogued as a product of intentional learning. The tenets of incidental learning (- attention, - intention, - awareness) (Rieder, 2003) are not met, since learners' attention was directed to the TWs through pre- and post-tasks (see 'Procedure' in Chapter 6); in other words, learners were aware of the nature of the intervention.

To conclude, it is worth remarking that vocabulary learning is influenced by word- and learner-related factors. A word's 'learning burden', defined as the amount of effort required to learn it (Swenson & West, 1934), will very much depend on the similarity between L1 and TL. Therefore, cognates will be better learned than non-cognates (d'Ydewalle and Van de Poel, 1999; Van der Slik, 2010; Vidal, 2011), and words that only include phonemes with a direct equivalent in the native language's phonological system will also be easier to learn than words with sounds not present in the L1 (Singleton, 1999). Nouns, rather than verbs and adverbs, are more readily learned than other parts of speech, primarily because they seem to be the most useful to encode and decode oral speech (Ellis, 1999) and they also evoke a clear mental image. Polysemous words also seem to be tougher to learn (Bensoussan & Laufer, 1984), both incidentally and intentionally, since inferring their meaning from context is much more difficult (learners need to disentangle which of the many meanings the word has is being used in that particular case). It is also worth noting that synonyms of already learned L2 words are said to be more difficult to acquire, as attaching more labels to the same mental representation or conceptual entity is a task that requires more mental effort and higher attentional resources (Singleton, 1999).

Eventually, another factor that needs to be mentioned is word frequency, that is, the number of occurrences of a word that learners encounter in the input. Many studies have explored the role that it plays in vocabulary learning and found a positive correlation between frequency of occurrence and word learning: the more encounters with the TWs, the better they are learned (Horst et al., 1998; Milton, 2008; Pellicer-Sánchez & Schmitt, 2010; Pigada & Schmitt, 2006; Rott, 1999; Uchihara, Webb, & Yanagisawa, 2019; van Zeeland & Schmitt, 2013a; Vidal, 2003, 2011; Waring & Takaki, 2003; Webb, 2007). This also holds true for studies exploring word frequency and vocabulary learning through multimodal input (Peters, Heynen, & Puimège, 2016; Peters & Webb, 2018; Rodgers, 2013). However, in the latter studies, the possibilities to learn words do not always raise proportionally depending on the number of encounters (Garnier, 2014), and significant differences in learning are found only between words appearing once or twice and items appearing very often (i.e., more than eight times) (Bisson, van Heuven, Conklin, & Tunney, 2014b).

Regarding learner-related factors, a higher language aptitude has been related with better recall and recognition of unknown TWs (Dahlen & Caldwell-Harris, 2013). Besides, scores on WM tests have been shown to significantly correlate with vocabulary scores on form recognition (Malone, 2018) and translation tests (Atkins & Baddeley, 1998). However, the most important learner-related factor seems to be proficiency level: research analysing its role on the learning of vocabulary through multimodal input has found many instances of the Matthew effect (Penno, Wilkinson, & Moore, 2002; Stanovich, 1986), or the rich-get-richer principle, according to which higher proficiency students tend to make greater language gains. Moreover, this seems to be applicable to different populations: pre-schoolers (Alexiou, 2015), high-schoolers (Aurstad, 2013; Kvitnes, 2013) or adults (Montero Perez et al., 2014; Peters & Webb, 2018), and different aspects of lexical knowledge: form recognition (Peters et al., 2016; Peters & Webb, 2018), form recall (Peters et al., 2016), meaning recognition (Peters et al., 2016; Peters & Webb, 2018) and meaning recall (Montero Perez, Peters, & Desmet, 2015; Peters & Webb, 2018).

4.2. Vocabulary acquisition from video viewing vs. no video viewing

Although this type of studies could have set the initial stage of research on vocabulary acquisition through multimodal input, only two instances of research comparing the effects of video viewing to no viewing for vocabulary learning (Peters & Webb, 2018; Rodgers, 2013) have been identified by the researcher, both of which analysing the behaviour of adult intermediate learners (see Table 4.1 at the end of this section for a summary of the studies). Rodgers (2013) is of particular importance for this doctoral dissertation since our study was partially inspired by Rodgers' design. His study was also taken as point of reference for some of the decisions made prior to and during the pedagogical intervention; for instance, the vocabulary post-task (see ‘Instruments’ section in Chapter 6) is very similar to Rodgers’ meaning recognition test.

Rodgers (2013) was one of the first longitudinal studies to thoroughly investigate incidental vocabulary acquisition through watching FL television. This dissertation was divided into five chapters covering: (1) comprehension of English-language TV, (2) incidental vocabulary learning, (3) effects of lexical coverage, (4) learners’ attitudes towards using television as a learning tool, and (5) the role of captions in all of the above. For relevance purposes, only studies (2) and (5) will be summarized in this chapter, and study (1) will be commented when analysing content comprehension. All experiments were conducted with the same TV series: *Chuck*, which was originally broadcasted in the US in 2007 and considered lexically-appropriate and less challenging than other authentic TV series available. Although episodes are quite long (average length: 43 minutes), a lexical analysis of the vocabulary used in the 10 episodes selected showed that 95% coverage was achieved at the 3,000-word level (including knowledge of marginal words and proper nouns). Moreover, the profile of this TV series was very similar to the analysis of a larger corpus of TV programmes conducted by Webb and Rodgers (2009b), so it was considered to be age- and level-appropriate and representative of other audiovisual input.

In Study (2), in order to investigate whether uncaptioned English-language television could benefit incidental vocabulary acquisition, Rodgers selected participants from nine university intact classes studying a variety of degrees. All were Japanese undergraduates who had been learning English for a minimum of seven years. According to standardized proficiency tests, their level varied from low-intermediate to intermediate. A final sample of 187 participants were allocated to the Experimental Group (EG), who saw the different episodes of the selected

TV series in the original uncaptioned version. 73 students were allocated to the CG, who just took the vocabulary pre- and post-tests without watching the TV series. The study took place over a period of 13 weekly sessions that included the viewing of 10 episodes plus two sessions devoted to the pre- and post-tests. At the beginning of the process, students in the EG took the Vocabulary Levels Test (Nation, 1983, 1990; Schmitt, Schmitt, & Clapham, 2001), a MC test in which they had to match words with their meanings. The original test contains ten sets of three words (30 words in total) in each of the following frequency bands: 2k, 3k, 5k, 10k and Academic Word List. In Rodgers' thesis, though, only 2k, 3k and 5k bands were administered to check participants' prior vocabulary knowledge. Regardless of the condition, participants took a pre-test consisting of 60 TWs specially devised according to the vocabulary used in the TV series. These 60 target items included single words appearing from 5 to 54 times in the whole series and all had a corpus frequency of between 3k and 10k, as analysed by the RANGE software (Nation & Heatley, 2002). The range of occurrence of these items was varied, since some of them appeared in all the 10 episodes selected while others appeared in only one.

The pre-test was carefully designed in a MC format with six options (key + 4 distractors + 'I don't know' option in last position) to test knowledge of the TWs. Two versions of the pre-test were created: sensitive (easy) and tough (difficult), the difference between the two being that MC options in the tough test shared some form and meaning aspects with the key, making it harder to discern between correct and incorrect options; this was not the case for the sensitive test. Both versions of the test were translated into students' mother tongue (i.e., Japanese) and they were accompanied by an audio file presenting the aural form of the TWs at the same time participants saw the written form on the paper test. These same tests, but in a different randomized order, were later administered as tough and sensitive post-tests.

Participants in the EG gained an average of six words, while those in the CG gained five. In addition, relative gains (taking into account the number of TWs already known on the pre-test) were also calculated and they showed that learners in the EG gained a 23% in the tough test and almost a 30% in the sensitive, whereas participants in the CG gained a 21% in the tough test and 25% in the sensitive, with a huge variation between minimum and maximum scores, as well as standard deviations. These results led to significant differences between both test versions (tough vs. sensitive) ($p<.001$) in the two groups. When comparing the two conditions, statistical tests also revealed a small but significant difference on incidental vocabulary acquisition in both tests ($p<.01$), indicating that being exposed to English-language television was beneficial for the acquisition of the target items. Nevertheless, a series of Pearson correlations showed that greater vocabulary knowledge, as measured by the Vocabulary Levels Test, did not result in greater relative gains (as far as the EG is concerned). In a separate analysis, frequency of occurrence proved to be significant in the tough test only, whereas range (the distribution of such repetitions) did not turn out to be a significant factor. However, when the two (frequency and range) were combined into a single measurement, a small but significant correlation with relative gains was found. Therefore, results indicate that "learning is more apt to take place when vocabulary is encountered more often in a single episode" (p. 96). Rodgers accounts for these positive results by stating that presenting contextualized words increases learners' sense of a word's meaning and usage. At the same time, he points out that gains from initial episodes may have been lost at the time of testing and that partial knowledge of non-tested words may have been also increased. He concludes that viewing successive episodes of a single television series rises the opportunities for and the chances of incidental vocabulary acquisition.

As we have seen, Rodgers first analysed whether FL television viewing had an impact on vocabulary uptake. However, it remained unknown whether textual support in the form of captions would enhance this vocabulary learning even more. That is why, in Study (5), Rodgers compared the effects of test taking on vocabulary acquisition (the CG in Study 2) and captioned English-language television viewing. To do so, an additional group watching captioned TV series for the same period of time was selected, and the performance of 40 university freshman students with a (pre-)intermediate level of English was compared to that of students in the CG in Study (2). The same procedure followed in the previous study was adopted: students watched a total of 10 episodes with English captions and also took the modified version of the Vocabulary Levels Test, plus the two vocabulary tests (tough and sensitive) at the beginning and end of the experiment. Results showed that, when learners watched the videos with captions, they gained an average of 6 words regardless of the test's sensitivity, which converts into a relative gain of 24% on the tough test and 29% on the sensitive. Surprisingly, these relative gains were not significantly larger than those of the CG taking the same tests. The author then concluded that captions were not as beneficial as they were thought to be. Moreover, contrary to what was observed in the non-captioned study, higher scores on the VS test led to more relative gains in the captioned group. As in Study (2), relative gains and frequency of occurrence correlated at a significant level in the hardest version of the test, although this was not the case for range. In the sensitive test, range proved to be negatively correlated with learning; and when a combination of the two factors was analysed, significant results were obtained on both tests.

With these two studies, Rodgers concluded that uncaptioned FL television was positive for the acquisition of vocabulary, and he also claimed that captioned television was not so beneficial. However, he could not yet compare captioned and uncaptioned television viewing since the baseline for comparison in both studies was the CG, who only took the vocabulary tests. To fill this gap, Rodgers (2013) also compared the performance of those participants who were exposed to captioned television viewing and those watching the TV series without textual support[7]. In other words, the EG in Study (2) served as baseline for this comparison and became a CG. Results did not vary extensively between both conditions on the two tests. On the tough version, the captions group gained 6.03 words whereas the non-captioned gained 6.36 words. A similar tendency was observed in the easier version of the test, since the no-captions group (6.78) gained more words than the captions counterpart (5.93). As expected, statistical analyses revealed no significant differences between captions and no-captions groups (p=.494 on the tough test and p=.159 on the sensitive). The same tendency was observed when relative gains were the object of analysis, concluding that the presence of captions did not have an effect on participants' vocabulary gains. Rodgers attributes this lack of statistical difference between the captioned and the uncaptioned conditions to the greater spacing between TW encounters. As previous research used briefer videos, TWs appeared in a relatively shorter period of time and, thus, it may have been easier to focus on them and retrieve the words later on. In longitudinal studies, this becomes more difficult, as there is a higher concentration of unknown vocabulary which also needs to be processed. Nevertheless, the researcher highlights that captions are not detrimental to learning.

Overall, results from Rodgers (2013) showed a positive effect of watching ten episodes of an American TV series on vocabulary acquisition. These results can be extended to smaller amounts of television viewing (e.g., one-hour long documentary) (Peter & Webb, 2018). This more recent study aimed at investigating the incidental learning of new lexical items through

[7] This study in Rodgers (2013) should be introduced later in section 4.3., as it compares subtitled and non-subtitled video viewing. However, in order not to split Rodgers' thesis across different sections, it was decided to present it in section 4.2.

viewing a full-length television programme and it also focused on the influence of some factors (namely, frequency of occurrence, prior vocabulary knowledge, cognateness and word relevance) on the acquisition of such items. Two cohorts of Flemish EFL learners (one for Experiment 1 and the other for Experiment 2), with a (low-)intermediate level of the TL, were selected to take part in this project. In the two experiments, half of the students were allocated to the EG, who was exposed to a non-abridged one-hour long BBC documentary with no textual support, while the rest of the learners were assigned to the CG, with no exposure to audio-visual input (they only took the tests). Learners were tested on three measures of vocabulary knowledge (although only two were finally analysed): meaning recall (Experiment 1) and meaning recognition (Experiment 2). One week before the beginning of the treatment, learners were pre-tested on 64 TWs (which had been thoroughly scrutinized in a pilot study), varying in frequency (1-6 times), degree of relevance and cognateness; they also took a general vocabulary knowledge test. One week later, some learners watched the video and all of them took the immediate post-tests, followed by the delayed post-tests one week after the end of the experiment. Absolute and relative gains were computed, and statistical analyses showed that those in the EG learned more TWs than learners who were not exposed to video (both meaning recall and recognition were significant). Delayed post-tests did not end up being a reliable measure since deliberate learning took place between the two testing times. In the two aspects of vocabulary knowledge analysed, a positive relationship between learners' prior vocabulary knowledge and learning was observed, although cognateness was the parameter with the largest effect size (a cognate had between 2.5 and 8 times more chances of being learned than a non-cognate). Frequency was also shown to be a parameter which positively correlated with learning, but word relevance (i.e., the degree to which TWs were useful for understanding the documentary) did not play a significant role.

Several conclusions can be drawn from this study: first, that video exposure is beneficial for incidental vocabulary acquisition since the EG outperformed the CG in the two tests analysed. Furthermore, it also indicates that prior vocabulary knowledge (or proficiency level), frequency of occurrence and cognateness are probably the three most influential parameters determining word learning from video viewing. However, the authors advocate for further research on the topic as results cannot be extrapolated to other EFL populations due to the strong presence of English in Flanders, where television and movies are not dubbed (European Commission, 2011).

It is important to bear in mind that, in the two studies summarised in this section, the CG did not undergo any kind of treatment and did not receive any input involving the target vocabulary; they only took the tests. Hence, it should be noted as well that more research comparing video viewing with other teaching techniques ought to be conducted to draw clearer conclusions on the impact of such modality of input in FL acquisition. That said, there have been some studies aiming at comparing captioned and uncaptioned television viewing, in line with one of the studies in Rodgers (2013). In such experiments, the specific effects of captions on word learning can be more easily determined. This body of research is presented in the next section of this review.

Table 4.1
Summary of the studies presented in section 4.2. on vocabulary acquisition from video viewing vs. no video viewing.

Study	Participants	Experimental design	Input	Vocabulary test	Results
Rodgers (2013)	(Pre-)intermediate Japanese undergraduates	Viewing uncaptioned FL TV series vs. test taking	10 episodes of an American TV series (≈ 43' each)	60 TWs, MC meaning recognition test with two sensitivities	FL television beneficial for vocabulary acquisition
		Viewing captioned FL TV series vs. test taking			No significant difference between experimental conditions
		Viewing captioned vs. uncaptioned FL TV series			*No significant difference between captioned and uncaptioned television*
Peters & Webb (2018)	(Low-)intermediate Flemish EFL learners	Viewing uncaptioned television vs. test taking	One-hour long documentary	64 TWs, meaning recall and meaning recognition vocabulary tests	Viewing the documentary beneficial for vocabulary acquisition

4.3. Vocabulary acquisition from subtitled vs. non-subtitled video viewing

The studies presented below, all comparing the effects of video viewing with and without textual support, will be introduced according to the participants' proficiency level (advanced, intermediate and beginner) and, at their turn, according to the results obtained –more or less benefits of subtitling– (see Table 4.2 at the end of the section for a summary of all these studies).

4.3.1. Advanced learners.

The most relevant studies with advanced students are summarized in first place. Most of them have been conducted with undergraduate students at the same learning institutions where the research was carried out. To start with, the benefits of subtitling for an advanced population do not seem to be restricted to written vocabulary, since some studies have found it to be beneficial for oral word recognition too. In this respect, Markham (1999) analysed the effects of captioned videos on advanced ESL learners' listening word recognition. In order to answer the RQs, he pooled 118 university learners from two different classes (most of them from Asian descent and aged 17-21). They were exposed to two videos about very different topics (marine biology vs. civil rights). For each of the videos, a 50-item MC listening test was created in which participants listened to intact sentences taken from the videos and to four possible answers including the key; the three other distractors were selected on the basis of their phonological similarity with the correct answer. Each class saw the videos either with or without captions and they completed the corresponding listening word test after viewing them. It was found that "the availability of captions did significantly improve the ESL students' listening ability to recognize words on the videotapes that also appeared on the subsequent listening-only multiple-choice test" (p. 324). This effect was consistent regardless of the passage content, revealing a beneficial aspect of multimodal input for word recognition and emphasizing the potential value of using captions for a variety of L2 pedagogical purposes.

Following a much more comprehensive approach and taking into account language proficiency and language distance, Winke et al. (2010) investigated the role that captions play in video-based listening activities. 150 second- and fourth-year university learners of Spanish and Russian, as well as second-year learners of Arabic and Chinese, volunteered for the study, resulting in six different groups. Three short excerpts of English documentaries were chosen as input video materials and they were edited and translated into the four TLs. To these videos, captions were added resulting in 12 captioned and 12 uncaptioned videos. Prior to the beginning of the experiment, learners were tested on their depth of knowledge of the TWs using the Vocabulary Knowledge Scale (VKS) (Wesche & Paribakht, 1996); afterwards, they watched each of the three clips twice, once with captions and once without them, except for the Spanish group, in which some students saw the clips twice without captions or twice with them. For each of the four TLs, half of the learners saw the captioned video first and the other half saw it in second place. After watching each clip, learners were given a vocabulary test in which half of the keywords were presented aurally and the other half in written, and they were asked to translate them into English. This was followed by a comprehension test. The authors conclude that different modalities of input facilitate vocabulary acquisition, as those who were exposed to captions twice significantly outperformed the other groups on the two tests administered (with written and aural input). Besides, presenting the video first with captions allowed learners to isolate what they perceived to be important and helped them to focus their attention on that particular piece of information (see also Montero Perez et al., 2014). Finally, proficiency also played a significant role regarding the scores on both tests (written and aural input), although it did not affect the benefits derived from captions order. The authors suggest that “captions are beneficial for a range of proficiency levels, as long as the videos are matched appropriately in terms of content and complexity (not too hard *and* not too easy) to the proficiency level of the language learners” (p. 81).

Nevertheless, not all studies with an advanced population found a positive association between audiovisual input and vocabulary learning. A couple of investigations (Etemadi, 2012; Nagira, 2011) did not find any difference between subtitled and unsubtitled conditions although none of them reported a negative effect for subtitling video viewing either. Etemadi (2012) explored the benefits of exposing two groups of English translation majors, having had at least six years of formal instruction, to the same two clips in two different conditions: bimodal subtitling and unsubtitled video. After watching each of the clips, they were given 10 MC questions tapping on their knowledge of infrequent English words. Descriptive results showed that when clips were watched in the subtitling mode, a higher vocabulary retention was possible, but the difference between groups was not significant. Nagira (2011) used Japanese undergraduate and postgraduate students to explore the effects of watching two short videos with bimodal subtitles (in English) or without textual support at three testing times (pre-, post- and delayed post-test). After analysing the three VKS, an increase over time (from pre- to post-test) was observed in the two groups, although significant differences between the two conditions were only identified for one of the two videos. Moreover, little decay was observed in the delayed post-test one week after the end of the treatment, irrespective of the condition. She hypothesizes that the slight advantage of the captioned condition may be because captions help students to distinguish aurally presented words when they may have otherwise failed to make the adjustment between aural sound and written form.

As can be seen in the studies summarized so far, subtitles have been found to be mostly positive for the acquisition of new vocabulary. In some of the cases, it was statistically proven that being exposed to subtitled videos did make a difference in the acquisition of new lexical terms in comparison to viewing videos without subtitles. In some others, the difference was still observable at the descriptive level, but it was not ratified by inferential statistical tests. It could be argued that this practice was not negative either, since uncaptioned video viewing was not found to be significantly more beneficial than captioned video. To the best of our knowledge, only one study has been identified in which the non-captioned condition was significantly better than the subtitled one: Hsu (2013) investigated the percentage of low-frequency productive vocabulary in students' writing after watching a video clip closely connected to the paragraph they were required to write. 50 advanced learners studying English at a Taiwanese university were recruited and each of them was exposed to videos in four distinct conditions: captioned video, non-captioned video, video and captions only (no audio), and an audio-only condition. All videos lasted no more than 9 minutes and were suitable for participants' proficiency level. Students also took a pre-test writing task, which was later used as baseline to analyse the increase in the usage of low-frequency vocabulary after having watched the clips. Compositions were analysed using the RANGE software (Nation & Heatley, 2002) focusing on words beyond the 2k level, which are considered to be mid- and low-frequency vocabulary (Nation, 2013). All participants, regardless of the condition, improved in their use of low-frequency free active vocabulary after being exposed to any modality of video input. Nevertheless, when results were scrutinized according to the different modalities, it was shown that the highest proportion of low-frequency productive vocabulary was achieved when participants watched the clip without subtitles, followed by captioned video, audio-only and no-audio conditions. Therefore, introducing video support increased the use of infrequent vocabulary in students' writing, but adding extra textual support did not result in a better performance. The study emphasizes the use of videos as a tool to improve learner's productive vocabulary, showing that not only can it help to identify unknown vocabulary, but also use it in free production tasks.

In sum, research conducted with advanced learners, mainly university students, has pointed out subtle differences in the learning of vocabulary through subtitled videos. Most studies showed the benefits of subtitles in the learning of lexical items (at least at the receptive level), as the EGs significantly outperformed the others (namely, the non-captioned conditions). Nevertheless, the studies conducted were one-shot studies, with videos lasting no more than a few minutes, and little is known about the effects of repeated video exposure. As the present dissertation seeks to explore the effects of multimodal input on learners at different proficiency levels, the most relevant research with intermediate and beginner learners will be summarized in the following sections.

4.3.2. Intermediate learners.

Several studies on subtitling and vocabulary acquisition have been conducted with intermediate learners and the most relevant studies with this population will be reviewed in this section. To begin with, d'Ydewalle and Pavakanun (1997) conducted one of the first large-scale studies with high-school students (aged 16-17): it followed a three per three per nine approach (81 conditions in total), in which nine FLs of different typology were combined with three different types of soundtrack (Dutch -L1-, FL, and no soundtrack) and three subtitling modes (Dutch, FL, and no subtitles). After watching a 15-minute video, students were administered a battery of tests including a meaning recognition vocabulary test with those words appearing five times or more in the clip, a sentence construction test, identification tests and a simple structural test, in which learners had to choose the right form of a sentence. These four tests were matched for the modality of input learners had been exposed to, either auditory (if they were exposed to soundtrack only) or written (in case they were exposed to subtitled video). As no pre-test was administered, the performance in those conditions in which the FL was not presented was taken as baseline. It was shown that language distance and typological similarity did not determine language acquisition through subtitling video. Interestingly, reversed subtitling proved to be the best condition to foster learning in the four tests administered, although the bimodal subtitling condition also led to positive results, as participants seem to “connect the pictorial discourse and the meaning of the spoken foreign language” (p. 153). The authors affirm that language acquisition is possible simply by watching a short subtitled movie.

Although d’Ydewalle and Pavakanun (1997) found that reversed subtitling (L2 text and L1 audio) led to higher levels of vocabulary knowledge, Baltova (1999) challenged this idea, in what was probably the first doctoral dissertation on the learning of vocabulary through video viewing with on-screen text. In this research, a complex design with reversed, bimodal and unsubtitled (hereafter traditional) conditions was adopted and different video sequences were used. 93 Grade 11 students from six schools in Toronto’s metropolitan area studying core French as an L2 were recruited for the study. Out of them, 30 were allocated to the traditional condition, 34 to the bimodal condition and 29 to the reversed condition. After carefully examining the sample of participants, the author concluded that learners across the different experimental conditions were comparable.

As stated, there were three experimental ‘combinations’; in all of them participants watched a seven-minute video a total of three times. Students in the ‘reversed experimental condition’ first saw the video with English audio and French subtitles, followed by a bimodal presentation

(French audio and subtitles) and then a traditional / control presentation, in which the audio was in the L2 and no on-screen text was added. In the 'bimodal experimental condition', learners saw the video in a bimodal format two times in a row, followed by a traditional presentation. Finally, students in the 'traditional condition' saw the same video with French audio and no subtitles three consecutive times.

Data collection took three days in which learners were pre-tested on the first. On the second, they saw the videos and took immediate tests. Finally, two weeks later, they took the delayed post-tests. A total of five tests were administered: (1) a vocabulary recognition pre-test, based on Meara (1994), in which learners had to cross out the words they were not familiar with (no actual proof of knowledge was needed, though), (2) a VKS in which learners had to prove they knew the same TW meanings shown in the video, (3) a 30 TW C-Cloze test, (4) a content comprehension test (which will be explained in the next chapter), and (5) a general proficiency dictation test. Test (2) was given as a pre- and delayed post-test whereas test (3) was conceived as both immediate and delayed post-test, while students took the dictation test just at the beginning of the experiment to assess their proficiency level.

In order to check whether vocabulary acquisition and retention was greater in any of the experimental conditions, the results of the C-Cloze and VKS tests were taken into account, using the scores on the proficiency test as a covariate. As expected, learners in the two subtitled conditions acquired more vocabulary than the CG, although they also showed lesser levels of retention. Within the two subtitled conditions, the bimodal group significantly outperformed the reversed condition and this difference was maintained on the delayed test. Focusing on the VKS, a between-groups significance was only observed when comparing the bimodal group and the other two conditions. Overall, it could be stated that the bimodal condition led to the highest levels of learning, with reversed and traditional conditions being conducive to a poorer, but still comparable, performance. Finally, it is also noteworthy that proficiency (operationalized as participants' scores on the dictation test) explained a significant proportion of the variance in vocabulary learning and retention.

Delayed post-tests also evinced that watching videos with on-screen text leads to long-term retention of lexical terms, besides immediate recall. Furthermore, although learning in the traditional condition was limited to a few words, Baltova (1999) claims that this proves the usefulness of "video as a successful medium for the teaching of new L2 vocabulary" (pp. 127-128). Other aspects that are worth mentioning is that saliency was observed to be a relevant factor in the acquisition of vocabulary, in the same way as the presence of the FL in the subtitles. Those in the reversed condition were exposed to French subtitles only once and did poorly than those in the bimodal condition, who had been exposed to French subtitles twice, (this allowed them to make form-meaning connections faster and more efficiently). As a conclusion, the author advocates for a greater use of authentic audio-visual materials in the language classroom, bearing in mind they must be age- and level-appropriate. A call for research on long-term effects and on the nature of TWs (in terms of concreteness, cognateness, frequency of occurrence, Part of Speech –PoS–, etc.) is also made.

The two studies presented above explored the efficacy of bimodal or reversed subtitling on the acquisition of FL vocabulary. However, they did not include other subtitling or captioning techniques such as keywords or highlighted keywords. This gap in research was precisely filled by Montero Perez et al. (2014). They analysed the effects of full captions, keyword captions and highlighted keyword captions on participants' content comprehension and incidental vocabulary acquisition; the study also included a non-captioned group. 133 undergraduate

Dutch speakers from Belgian origin studying law at a Flemish university were selected as participants and assigned to one of the four conditions previously mentioned. The authors selected three short videos (2-4 minutes) and 17 TWs (prior knowledge of TWs with a group of similar proficiency to that of participants was assessed). These 17 words appeared highlighted in the highlighted keyword condition and in isolation in the keyword captioning condition, which only presented the 17% of the full script (this percentage was considered essential to understand the clips). One month before watching the videos, students took a meaning recall test following the VKS format. After the experimental phase (i.e., video viewing), participants took a total of four vocabulary tests: form recognition, in which students had to say whether or not they had seen the word in the clips; clip association, where they had to decide in which clip the word had appeared; the same meaning recall test taken at the beginning of the intervention; and a meaning recognition test, where learners had to select the best Dutch translation for the French word.

Results showed that all vocabulary tests scores correlated at a significant level and that type of captioning proved to be a determinant factor explaining much of the variance in the scores. Furthermore, type of captioning significantly affected form recognition, clip association, and meaning recognition. Significant differences were found between the no captioning condition and the other three on form recognition and clip association. In relation to meaning recognition, only the keyword captioning and the full captioning with highlighted keywords conditions significantly outperformed the CG. No differences were observed at the meaning recall level. The authors emphasize the important role that VS played in all vocabulary aspects analysed and that captions help learners to decode and to isolate words, as seen in the differences on clip association and word recognition tests. In contrast, less differences were found on meaning recognition and recall, maybe due to the test’s demands (it could have been too difficult for students as they were expected to watch the clip, read the captions, infer the meaning of new words and remember some TW meanings). To finish, the authors also advocate for more longitudinal research on the effects of captions and with full-length TV programmes.

Montero Perez et al. (2018) went a step further and found that an innovative technique such as glossed keyword captioning was also beneficial for vocabulary acquisition. Besides, they also compared the extent to which test announcement hindered (or facilitated) vocabulary acquisition. 227 (high-)intermediate undergraduate learners of French aged 17-21 were recruited for the study. As in previous studies (see Montero Perez et al., 2015), the intentional learning group received explicit information about upcoming vocabulary post-tests, while the incidental learning group was not told in advance about such vocabulary post-tests. Participants were allocated to one of the four experimental conditions: no captioning, full captioning, keyword captioning, with the only presence of 17% of the script, and the innovative glossed keyword captions, thanks to which learners had the opportunity to access the meaning of the keywords. The same three short clips (with a maximum length of 5 minutes) that were part of previous studies were also used in this one.

18 TWs from different categories were included in the final design and participants were faced with a battery of tests: a VS test and several vocabulary post-tests (see Peters et al., 2016 or Montero Perez et al., 2014), tapping into form recognition, clip association, meaning recall (translation test) and meaning recognition. All these were accompanied by general comprehension questions for each clip to check whether students were attentive to the videos.

Inferential analyses revealed a significant main effect of type of captioning on all vocabulary tests, although no differences were observed between incidental and intentional conditions.

Captioning proved to be positive for vocabulary acquisition as the three captioning groups outperformed non-captioned learners on all tests except for the meaning recall, where only the glossed captioned group turned to be better than the others. Results confirmed the Matthew effect, as a positive relationship between learners' VS and vocabulary learning was observed. Besides, authors concluded that giving students the chance to immediately access the meaning of TWs resulted in an increase in the scores on the meaning recall test, facilitating the consolidation of form-meaning connections.

As was the case with advanced learners, not all studies with intermediate learners have come to the same positive conclusions. For instance, Yuksel and Tanriverdi (2009) found no differences between subtitled and unsubtitled conditions in 120 intermediate students, half of whom saw the video with textual support while the others did not. In this case, the VKS was administered to students before and after watching the video (an extract from a *Seinfeld* episode lasting nine minutes). Again, and in line with other studies (e.g., Aurstad, 2013; d'Ydewalle & Pavakanun 1995, 1997), improvement was observed from the beginning to the end of the treatment, although no differences were found between the two conditions. The authors argue that video viewing actually facilitated vocabulary acquisition (and not captions themselves) as it presented words in context. However, it must be borne in mind that TWs did not appear frequently and that learners also had to build knowledge on who the characters were and the relationship between them, which may not have left much time and resources to focus on unknown words.

Using exactly the same design, similar results were observed by Karakas and Saricoban (2012), who investigated whether subtitles facilitated the acquisition of vocabulary in animated cartoons. They exposed 42 upper-intermediate Turkish learners studying English at university to two episodes of a very popular animated cartoon originally broadcasted in the US. A final selection of 18 TWs was included in a self-reported VKS test one week before and one week after watching the episode. A series of *t*-tests were conducted to evaluate any progress between pre- and post-test as well as any difference between conditions. Although both groups experienced an improvement between the beginning and the end of the experiment, differences between groups were not large enough to reach statistical significance, even though the group exposed to subtitles obtained higher means than the CG at the end of the treatment. In conclusion, the study further supports the beneficial effects of exposure to multimodal input for incidental vocabulary learning (but textual support was not considered necessary).

If the two studies presented above analysed the behaviour of intermediate university learners, two MA theses conducted in the Norwegian school system selected high-schoolers (aged 16 in the first and 17 in the second) and found similar results. Kvitnes (2013) and Aurstad (2013) investigated long-term effects of video viewing on vocabulary acquisition by administering a lexical decision and a word definition task. In the two cases, three conditions were compared: either watching the FL video with English subtitles, Norwegian subtitles or without any textual support. After watching a 20-minute episode of a well-known American animated cartoon, students were given a comprehension test in the form of MC questions and, four weeks later, a lexical decision task, in which they were asked to identify which words had appeared in the episode, and a meaning recognition task, where they had to pick up the correct translation of TWs and idioms out of four different possibilities. However, no previous knowledge of these TWs and phrases was checked before the beginning of the experiment (only a general vocabulary test was administered as part of a larger proficiency test).

In the first group of adolescents (Kvitnes, 2013), content comprehension of the episode was an important predictor of learners' scores in the word definition task administered four weeks later (p=.013). Focusing on both vocabulary tests, condition was not a predictor for the scores because participants from the three groups obtained very similar results. Overall, it seems that having the visual and auditory representations of the lexical item does not facilitate memorization of TWs four weeks after having seen them, neither it facilitates to a great extent the recognition of their meanings. Although EGs scored a 10% higher than the CG, condition was not a significant factor in the final statistical model.

In the study with forty-nine 17-year-olds (Aurstad, 2013), a similar pattern was found in the word definition task, since, once again, the difference between experimental conditions was not statistically significant. However, grammatical knowledge, among others, proved to be a good predictor of learners' vocabulary scores. Similarly, the results of the lexical decision task did not differ across groups, mirroring the behaviour observed with participants from the younger age group.

From these two MA theses, the authors conclude that, after watching a single episode of an animated cartoon, the benefits of multimodal input are hardly observable after four weeks. It is believed that, rather than subtitles, general proficiency and grammar knowledge are the two factors that have a stronger influence on the learning of new vocabulary. The answer to whether subtitles have positive long-term effects is something that authors leave for further research.

In brief, not all research conducted so far including the experimental condition of watching the videos without subtitles has drawn the same conclusions for intermediate learners. Some studies did find a positive effect of textual support, while others spotted no significant differences between watching the videos with or without textual support. However, on-screen text was not found to be negative for language learning either. Besides, other subtitling techniques (e.g., keyword or glossed keyword captioning) proved to be slightly more effective than more traditional approaches such as full captioning.

4.3.3. Beginner learners.

Regarding beginner learners, a very important distinction between proficiency level and age should be made, since the population that these studies targeted was quite heterogeneous and ranged from children in primary school (or kindergarten) to adults beginning to study a new FL later in life. One of the difficulties in studies with children is that they do not have their reading abilities fully developed and therefore it is problematic to conduct studies on subtitling with them. This is the main reason why studies with very young learners have tried to look from other perspectives at the effects of watching videos on language learning.

For instance, Rice, Hustin, Truglio, and Wright (1990) investigated vocabulary acquisition in children watching *Sesame Street* at home, in a less controlled environment. One-week TV viewing diaries were collected from children's parents every six months over a period of two years. One of the groups of the study (n=160) was followed from ages three to five and the other one (n=166) from five to seven. Together with the viewing diaries, other background measures were also considered (e.g., parental educational level, occupational status, presence of siblings) and the Peabody Picture Vocabulary Test (PPVT) (second edition) (Dunn & Dunn, 1981) was administered to all children before and after the data collection process. Several regressions showed a positive effect of watching the series in children aged 3-5, although this positive influence declined in children aged 5-7. No relationship was found between watching any other kind of programme and children's vocabulary scores. However, the authors claim that watching *Sesame Street* contributes to children's vocabulary development and they advocate for its use in preschool, as the study shows that incorporation of new words derived from viewing is a real possibility. As referred to in Chapter 3, this idea was also shared by Neuman and Koskinen (1992), when they analysed whether TV influenced the acquisition of science terms and concluded that multimodal input is a good way to foster language acquisition and literacy.

In the same line, Kuppens (2010) also tried to establish a set of relationships between informal exposure to TV and language skills (in this case, measured by translation tests). More specifically, the self-reported use of English language media by 374 primary school students and their performance on two translation tests (L1 to L2 and vice versa) were compared. Students were asked to fill in a survey with questions related to how much subtitled and unsubtitled TV programmes and movies they watched, and to their exposure to English songs, websites, radio, or computer games, among others. Interestingly, their scores on both translation tests were directly proportional to the number of hours spent watching subtitled English television, explaining a 25.5% of the variance on average and with a higher influence on girls than boys. Although the study has some limitations because 'proper language acquisition' could not be measured, it clearly shows how important extramural exposure to English television is and proves that children and low-level learners can benefit from subtitled TV viewing.

The benefits of informal TV viewing in beginner learners hinted at by the previous two studies were to a certain extent challenged by Alloway, Williams, Jones, and Cochrane (2014), who investigated the relationship between exposure to television and vocabulary skills in toddlers aged between two and three. They used a receptive picture vocabulary test (PPVT) and a carefully designed questionnaire, which asked parents about their children's television watching habits and about other literacy skills such as frequency of reading. According to the results they obtained, television watching did not have either a positive or negative effect on children's vocabulary scores in comparison to the benefits observed with reading educational books or being read stories (which explained 15.8% of the variance in participants' vocabulary scores). The authors argue that the programmes children were exposed to targeted older children and they may use language which toddlers are not ready to comprehend. Although watching TV did not impact negatively on toddlers' scores, they suggest that more beneficial activities, such as those developing short-term memory skills, should be planned and implemented if children's vocabulary abilities are to be fostered.

The positive effects of watching television for beginner learners are not restricted to informal contexts but have also been observed in more experimental settings. To start with, Alexiou (2015) conducted a study with 30 Greek EFL learners aged 4 to 6 (pooled from a nursery school in Greece) and studied their vocabulary uptake from watching five different episodes of a very well-known animated cartoon for pre-schoolers. Episodes were shown in an English-only condition and were not accompanied by any kind of explicit instruction. After watching each clip four items, an aural form recognition test was administered, in which students had to select the picture that best represented the word the researcher said out loud; being it closely similar in format to the PPVT (Dunn & Dunn, 2007). It was seen that pre-schoolers could correctly identify a third (i.e., 7.66) of the 21 TWs. Besides, age was also considered a determinant factor in the uptake of vocabulary, since older students consistently outperformed younger ones. It was also shown that cognates, as well as concrete and imaginable nouns and adjectives, were easier to pick up than abstract verbs and adverbs. Finally, frequency of occurrence was found to have no effect at this age: participants picked up some of the words that occurred less often and could miss some of those which were very salient and occurred many times. However, all these results need to be treated with caution since no previous knowledge of the content vocabulary was checked and there was not a control condition.

Positive results in experimental settings have also been observed in older populations with a beginner level of the TL. In a much-cited study, Koolstra and Beentjes (1999) explored the extent to which Dutch children aged 10 and 12 learned the FL through watching subtitled TV. A total of 246 students from Grades 4 and 6 were selected from three schools in the Netherlands; at each level they were divided into three groups and were matched for their level of English (measured with a picture matching vocabulary test). All watched a short TV programme: (i) a 15-minute documentary in English with Dutch subtitles, (ii) the same documentary in English without subtitles or (iii) a different documentary in Dutch without subtitles. After the treatment, children took a MC translation test asking for the Dutch equivalent of 35 English TWs. In addition, the two groups exposed to English audio were given a word recognition test. As in previous studies, results showed a significant effect of condition in the two tests administered, with learners who watched the video with subtitles scoring significantly higher than the others. As expected, sixth graders outperformed fourth graders.

Finally, d'Ydewalle and Pavakanun (1995) conducted two studies targeting two distinct populations: adults (Experiment 1) and adolescents (Experiment 2). A three-per-three design combining the language of the audio and that of subtitles (mother tongue, L2 or no soundtrack / subtitles) was adopted. In Experiment 1, 90 adult English NSs with a beginner level of the TL were recruited for the study and were exposed to a 12-minute excerpt from a well-known animated cartoon presented in English or Dutch (TL) soundtrack and / or subtitles. After watching the clip in one of these conditions, they were given a battery of tests including sentence construction and meaning recognition vocabulary tests. Statistical analyses showed a significant main effect for subtitling, and not for dubbing (i.e., the language in which the audio of the video was presented), on the vocabulary test (not on the sentence construction test), implying that the inclusion of textual support (either in L1 or L2) was beneficial for vocabulary acquisition. In Experiment 2, the same procedure was followed, although this time researchers recruited a total of 99 Grade 9 students in an immersion school in Belgium. In this second experiment, no differences were observed in any of the two tests that participants completed. In terms of age (comparing adults with ninth graders), older learners outperformed younger ones on the meaning recognition vocabulary test but, at the same time, Grade 9 students showed a better performance on the sentence construction test, probably attributable to the higher proficiency level they had as a result of attending an immersion school. Overall, the authors concluded that the reversed subtitling condition (English audio and Dutch subtitles) led to more learning as "providing the dominant (English) language through the more transient auditory channel gives the subjects more time to match the message with the translated text in the subtitle" (p. 62). The potential of subtitled video in the language classroom is also emphasized.

The benefits of exposing beginner learners to video viewing are especially relevant when modality of input and test modality are matched (i.e., when the test is aural or written depending on whether or not learners have been exposed to captions), as shown by Sydorenko (2010). She asked 26 English beginner learners of Russian (their average age being 20 years) to watch a video under three different conditions: with audio and captions (VAC), with audio but no captions (VA) or with captions and no audio (VC). Learners were tested on their recognition of written and aural forms of words as well as on their translation abilities. Regardless of the condition participants were allocated to, they were all exposed to three very short clips (2-3 minutes) containing 28 unknown Russian TWs in total; videos were shown a total of two times each. After watching them, participants were given a comprehension test (although it was not analysed), a written and aural word recognition test, an L2-L1 translation test (aural and written) and a word knowledge test working as a measure of prior knowledge of the TWs. Finally, a questionnaire enquiring about participants' attitudes and strategies used was also distributed. As far as word recognition and translation tests are concerned, there was not a significant main effect for test type (i.e., auditory vs. written tests). However, when analysing the overall learning of words (i.e., combining written and aural vocabulary tests), results showed a significant main effect for modality of input on the vocabulary translation test, suggesting that "the VAC combination is more favourable than the VA combination for learning the meanings of new words" (p. 58). In summary, results supported the idea that word form recognition is enhanced when modality of input and test modality are matched, in line with what Bird and Williams (2002) found with bimodal input. Moreover, Sydorenko also concludes that performing three different tasks (watching, reading and listening) resulted in better scores than performing two (watching, and reading or listening), contradicting CLT redundancy principle. It also seemed that captions helped beginner learners more than audio, as the VC condition led to better results than the VA condition, although she calls for further research on the topic.

Nevertheless, not all studies found such positive results with beginners. For example, following a design inspired by this doctoral dissertation, Galimberti (2016) explored the effects of watching a 22-minute episode of a TV series with L1 subtitles, L2 subtitles or no subtitles. In this case, participants were 52 Italian students aged 12 learning English at primary school. Before watching the episode, all learners were asked to take a form and meaning recall vocabulary pre-test in which they listened to the set of TWs in English and had to provide their

English written form and L1 translation. This same test was administered immediately after watching the video too, as a post-test. Together with the recall test, a MC meaning recognition test was distributed: it consisted of the TWs in their L2 forms and five possible meanings, out of which participants had to choose the correct translation. Although within-subject comparisons from beginning to end of the intervention were statistically significant, only those watching the L2 subtitled TV series could recall a significantly higher number of TWs compared to the other experimental conditions. No differences were observed between L1 subtitles and no textual support on the recall test and between any of the three conditions on the meaning recognition test. Possibly, children may have learned more from the soundtrack than from subtitles, as opposed to what happens with adults (see d'Ydewalle & Van de Poel, 1999), according to the author's interpretation. Moreover, a poor reading performance and less developed L2 skills could explain the small gains (10% of TWs were learned on average) in vocabulary recall. However, she highlights the fact that vocabulary acquisition is possible just after viewing one episode of a TV series.

Also, when Raine (2012) investigated the effectiveness of subtitled authentic videos as a way to increase depth of vocabulary knowledge, he found similar results as those by Galimberti (2016). He selected 39 Japanese university students with a low level of English and assigned them to one of the four experimental conditions: English video with no subtitles, video with English or Japanese subtitles, and dual subtitles combining both languages. After selecting six TWs that appeared in the video, the VKS was used as a pre- and post-test. The video, which lasted for six minutes, was projected twice and statistical tests conducted after watching it revealed no differences between the four conditions although those exposed to dual subtitles were the ones that progressed the most. However, it should be noted that participants significantly differed in their scores on the proficiency test they took. Therefore, results should be taken with caution as it is difficult to say whether the progress observed is due to the treatment or to the existing differences in proficiency.

Finally, some studies did not directly look at the effects of video viewing but focused on hands-on activities involving videos. A clear example of this type of experiments was conducted by Lertola (2012). In this study, sixteen university participants with an A2 level of the TL were allocated to two different conditions: experimental and control. The EG performed a subtitling task consisting in the translation of the dialogue of a short video clip from Italian into English, whereas the CG did other activities related to the same video, although they did not involve the creation of subtitles. The experiment, which lasted for four weeks, was initiated with a first exposure to the video by both groups, although the CG did not see it subtitled. In the following three weeks, the EG was exposed to the video with captions and translated the transcript from their mother tongue to the TL. So as to measure vocabulary, a VKS test with 15 TWs was administered before the start of the treatment, immediately after finishing it and two weeks later. Results confirmed an improvement across the three testing times for both groups. However, when scores were analysed cross-sectionally, no differences were observed on the immediate post-test scores, even though they appeared on the delayed post-test, in favour of the EG. Consequently, the study suggests that explicit attention to subtitles and meaning-making activities do not improve vocabulary learning scores in the short run although negative effects were not observed either. The study does not yield clear results about the role that the activity of subtitling itself plays in vocabulary learning: although differences were found on the delayed post-test, they were not observed at any other point and a similar longitudinal progression was seen in both groups.

To sum up this subsection on beginner learners, a common finding of the studies presented is that vocabulary gains were rather small, no matter whether participants were children, high-school learners or older populations, or the type of study conducted. What is worth noting, though, is that a wide range of studies have dealt with the topic in different ways: from experimental designs, such as Sydorenko (2010), to studies analysing TV viewing at home and vocabulary learning (e.g., Rice et al., 1990). Others have also explored which combination of audio and textual support could be conducive to more vocabulary learning (Galimberti, 2016) and results have been rather inconclusive, since some studies pointed towards a significant effect of subtitles on performance while others did not find any significant differences.

Table 4.2
Summary of the studies presented in section 4.3. on vocabulary acquisition from subtitled vs. non-subtitled video viewing.

Note. The dotted lines mark a difference in proficiency level (in descending order of proficiency: advanced, intermediate, beginner) of the participants in the studies.

Study	Participants	Experimental design	Input	Vocabulary test	Results
Markham (1999)	Adult university learners	Viewing captioned vs. uncaptioned FL video	Two short videos (12' and 13' each)	50 TWs, listening word recognition test	Captioning significantly improve listening word recognition
Winke et al. (2010)	2nd and 4th year university learners of Spanish, Russian, Arabic and Chinese		Three short clips (3'-5' each)	Vocabulary translation test with aural and written input	Captioning significantly beneficial for oral and written vocabulary tests
Etemadi (2012)	Iranian English translation majors	Viewing FL videos with vs. without bimodal subtitles	Two documentaries (20' and 30' each)	10 TWs, MC meaning recognition test	No difference between experimental conditions
Nagira (2011)	Japanese undergraduates and postgraduates		Two short clips (7' each)	VKS	Benefits of captioning limited to one clip
Hsu (2013)	Taiwanese advanced English majors	Effects of viewing captioned vs. uncaptioned FL video on productive vocabulary	Four short videos (≈ 8' each)	Free compositions based on the topic of the videos	Uncaptioned video promoted the use of less frequent vocabulary
d'Ydewalle & Pavakanun (1997)	Intermediate high-schoolers (aged 16-17)	Viewing video with audio and subtitles in the L1, L2, or no soundtrack or subtitles; 9 FLs involved	One short clip (15')	Aural and written meaning recognition vocabulary, sentence construction, simple identification, and structural tests	Reversed subtitling most beneficial; language distance and typology did not affect learning
Baltova (1999)	Grade 11 high-school learners	FL video viewing with L1, L2 or no subtitles	7-minute video (seen thrice)	Vocabulary recognition test, VKS and C-Cloze test	Subtitled video, esp. bimodal subtitling, led to more vocabulary learning

Study	Participants	Experimental design	Input	Vocabulary test	Results
Montero Perez et al. (2014)	(High-)intermediate Dutch undergraduates	FL video viewing with full, keyword, highlighted keyword or no captioning	Three short clips (2’-4’ each)	17 TWs, four vocabulary tests including form recognition, clip association, meaning recall and meaning recognition	Captioning always beneficial for form recognition and clip association, less beneficial for meaning recognition and no difference on meaning recall
Montero Perez et al. (2018)	(High-)intermediate undergraduate learners of French	Video viewing with full, keyword, glossed keyword and no captioning; comparing test announcement effect		18 TWs, four vocabulary tests including form recognition, clip association, meaning recall and meaning recognition	No difference if test announced. Captioning led to more vocabulary learning, esp. glossed keyword
Yuksel & Tanriverdi (2009)	Intermediate pre-university learners	Viewing captioned vs. uncaptioned FL video	Excerpt from a TV series (≈ 9’) (seen twice)	20 TWs, VKS	No difference between experimental conditions
Karacas & Saricoban (2012)	Upper-intermediate English majors		Two episodes of an animated cartoon	18 TWs, VKS	
Kvitnes (2013)	Intermediate high-school learners (aged 16)	FL video viewing with L1, L2 or no subtitles	One episode of an animated cartoon (20’)	Lexical decision and meaning recognition vocabulary tests	
Aurstad (2013)	Intermediate high-school learners (aged 17)				
Rice et al. (1990)	Children aged 3-5 and 5-7	Watching TV programmes at home	Episodes of *Sesame Street*	TV viewing diaries and PPVT	Positive effect of watching TV in children 3-5, not in children 5-7
Kuppens (2010)	Primary school students		Different types of TV programmes	L1-L2 and L2-L1 translation tests	Amount of subtitled TV viewing proportional to scores on translation tests

Study	Participants	Experimental design	Input	Vocabulary test	Results
Alloway et al. (2014)	Toddlers (aged 2-3)	Watching TV programmes at home	Different types of TV programmes	Receptive picture vocabulary test	No effects of TV viewing on vocabulary scores
Alexiou (2015)	Greek pre-school learners (aged 4-6)	Viewing FL video (no comparison)	Five episodes of an animated cartoon (≈5' each) (seen four times)	21 TWs, aural form recognition vocabulary test	Correct identification of a third of the TWs
Koolstra & Beentjes (1999)	Grade 4 and Grade 6 learners	Viewing subtitled vs. unsubtitled FL video	15-minute documentary	35 TWs, MC translation and recognition tests	Subtitled video viewing led to more vocabulary learning
d'Ydewalle & Pavakanun (1995)	Beginner university learners	Viewing video with audio and subtitles in the L1, L2, or no soundtrack or subtitles	12-minute excerpt from an animated cartoon	Sentence construction and meaning recognition vocabulary tests	Significant main effect of subtitling
	Grade 9 learners in an immersion school				No significant main effect of dubbing or subtitling
Sydorenko (2010)	Beginner university learners of Russian	Viewing video with audio and captions, only audio or only captions	Three short clips (2'-3' each) (seen twice)	28 TWs, written and aural word recognition test and L2-L1 translation test	Video with captions beneficial for learning word meaning
Galimberti (2016)	Grade 6 Italian learners	FL video viewing with L1, L2 or no subtitles	One episode of a TV series (22')	10 TWs, MC meaning recognition and form and meaning recall vocabulary tests	L2 subtitles positive for TW recall; no difference in the recognition test
Raine (2012)	Beginner Japanese university learners	FL video viewing with L1, L2, dual or no subtitles	One short video (6') (seen twice)	6 TWs, VKS	No difference between experimental conditions
Lertola (2012)	Beginner university learners	Translation of dialogue of a short video vs. traditional classroom activities	L1-L2 translation of dialogue of a short video	15 TWs, VKS (immediate and delayed)	No difference in immediate testing; difference in delayed testing

4.4. Vocabulary acquisition from different types of subtitles

There is also a line of research focusing on the effectiveness of different types of subtitling, namely by comparing full captioning and keyword captioning or by exploring the effect of enhancement techniques such as highlighting words or changing the size and font of the subtitles. Some have also tried to find the combination of text and sound (e.g., L1 text and L2 sound or L2 sound and L2 text) that is most beneficial for FL learning. In this section, then, and in contrast with the studies in the previous sections, all studies always used videos accompanied with textual support (see Table 4.3 for a summary of the studies reviewed in this section).

4.4.1. Advanced learners.

Within this line of research, there is only one study with advanced learners, Zarei (2009), which analysed the effect of different types of subtitles on L2 vocabulary recognition and recall in 90 Iranian undergraduates (ranging in age from 19 to 26 years). After completing a 100-item vocabulary test to check which words were already known before the experimental phase, participants saw nine episodes of a British TV series in one of these three conditions: English audio and subtitles (bimodal condition), English audio and Persian (L1) subtitles (standard subtitling) or Persian soundtrack and English text (reversed condition). Participants watched all the episodes and completed a vocabulary recognition post-test that included 40 words which were deemed to be unknown at the beginning of the experiment; besides, another vocabulary recall test (fill-in-the-blanks format) was used to gauge progress. Statistical analyses showed a significant effect of the subtitling condition on learners' vocabulary recognition, with the reversed subtitling condition being the least beneficial. In terms of vocabulary recall, a significant effect of the subtitling condition was also found, with bimodal subtitles being the most favourable method of presentation, followed by standard and reversed conditions (in this same order). According to the study, the subtitling condition explained 20% of the variance in vocabulary recognition whereas it accounted up to 26% when it came to vocabulary recall. A possible explanation for the poor performance of those participants exposed to reversed subtitles could be that they were presented with the soundtrack in their native language, and they may not have been pushed to read the subtitles. When this happens, it is difficult that they might end up learning new words, either productively or receptively. Although results would need to be confirmed by other studies, they can be taken as support of the fact that advanced students should be exposed to bimodal subtitling when trying to foster vocabulary recognition and recall.

4.4.2. Intermediate learners.

There is more research in the case of intermediate FL learners, and it has proven that bimodal subtitling can be the most beneficial condition for vocabulary learning. For instance, one of the most recent studies was conducted by Peters et al. (2016). In the two experiments of this study, the authors explored the effects of L1 subtitles and captions on form recognition and the form-meaning mapping in (low-)intermediate EFL learners in Belgium. In both experiments, one intact class was allocated to the L1 subtitled condition and another to a captioned condition. In Study 1, following a pre- / post-test design, 28 secondary school students were selected: sixteen 17-year-olds were exposed to a thirteen-minute clip subtitled in their L1 (Dutch) and twelve 18-year-old students watched the same clip but with L2 subtitles (i.e., captions). One week before they watched the clip, learners took a MC VS test with words from different frequency bands and a spoken form recognition and meaning recall test that asked them to state whether each of the 50 words they listened to (39 TWs + 11 distractors) had been heard or seen before and provide its meaning in case they knew it. These two same tests (excluding the VS test) were administered again immediately after watching the clip to measure any learning gains. A Generalized Estimating Equation showed that learners in the captioned group recognized more TWs than their peers in the L1 subtitled condition, although VS was one of the key factors explaining a high percentage of the variance. They conclude that L2 subtitles are more beneficial for learning formal aspects of word knowledge (i.e., form recognition). In the meaning recall test, regardless of the experimental condition, learners got a 20% of the answers right on average. Although the two groups scored similarly, TWs' frequency of occurrence and VS turned out to be two important parameters that significantly contributed to the statistical model.

The second experiment followed the same design as the first, the difference being that participants were selected from a vocational school. Moreover, learners were also tested on form recall, which is considered to be a more difficult task, and meaning recognition, which was in principle an easier test for participants. A total of 18 participants, ranging in age between 17 to 20, were recruited and assigned to the two experimental conditions (L1 vs. L2 subtitles). This time, an episode of a very popular American TV series was selected, and Dutch / English subtitles were added, from which 18 items varying in frequency were chosen as TWs. After doing a VS test two weeks prior to the beginning of the experiment, participants took three vocabulary pre-tests (form recall, form recognition and meaning recognition, in this same order). These were also administered immediately after watching the episode. The Generalized Estimating Equation model showed that captions were beneficial for learning the form of unknown words, with VS and frequency of occurrence playing a major role. In spite of this fact, type of subtitling was not a determinant factor on the form and meaning recognition tests.

After comparing and analysing the results of the two experiments, the authors conclude that captions benefit learning more than L1 subtitles. Contrary to what other studies found (e.g., Borrás & Lafayette, 1994), L1 subtitles did not facilitate the acquisition of meaning and researchers argued that "when provided with the meaning via L1 subtitles, learners also see the visual clue or image but it might be more difficult to link the meaning of the target item to the L2 aural form in the speech stream" (p. 145). Eventually, they also emphasized the important role played by VS, as it positively correlated with the learning of target items in the two experiments, and the beneficial, yet inconclusive, relationship between number of occurrences and learning.

Comparable results to those found by Peters et al. (2016) were also observed among university students by Zarei and Rashvand (2011). The study, which followed a close design to Zarei (2009), included a larger number of experimental conditions. This time, the authors investigated intra and interlingual subtitles as well as the effect of verbatim (exact transcription of oral materials) versus non-verbatim subtitles (slight adaptation of the oral conversation by eliminating redundant information and some language particles in order to facilitate comprehension). In this study, 120 adult university students with an intermediate level of the TL and from Iranian descent were recruited. They were divided into the four possible conditions: verbatim intralingual (English audio and text; that is, bimodal subtitling), verbatim interlingual (English audio and L1 subtitles; that is, standard subtitling), non-verbatim intralingual and non-verbatim interlingual. An authentic but non-original American comedy lasting 70 minutes was selected to be watched. Prior to watching the movie, students took a general proficiency test and a 100-item pre-test to check their knowledge of the words used in the film. After watching the movie, they were asked to write a one-page summary (to make sure they understood it) and to take two post-tests: a MC vocabulary recognition task and a fill-in-the-blank production test. No significant differences were found between intra and interlingual subtitles on vocabulary recognition, although a difference was found between verbatim and non-verbatim text, the latter being better for FL learning. In contrast, as far as vocabulary production is concerned, whether subtitles had been adapted or not (i.e., verbatim or non-verbatim) did not prove to be statistically significant, whereas the combination of audio and textual support caused significant differences: the intralingual combination was more beneficial for learners, mainly due to the easier access to the written forms of words, which can be conducive to TW production. Non-verbatim's superiority could be explained by the fact that learners' attention is directed towards the needed information, and students are also less distracted by redundant information that does not need further processing.

The benefits of bimodal subtitling have also been found to apply to colloquial language learning, even though they seem to be restricted to long-term testing, as shown by Frumuselu's doctoral dissertation. A total of four studies were conducted, although only three will be reported here since the fourth one is not directly linked to the present dissertation. Her thesis aimed at investigating the short- and long-term effects of captioned and subtitled video upon informal and colloquial language acquisition. English majors were pooled from second-year university classes, varying widely in proficiency (from A2 to C1 levels according to the Common European Framework of Reference for languages –CEFR–) and in age (from 19 to 25; median: 20 years old). A total of 49 learners participated in the different studies, although not all of them were included in the three experiments. 90% of them were Catalan / Spanish bilinguals and the other 10% spoke a wide variety of languages, none of them being English, which was the TL for all studies. All the studies were based on a selection of (unconnected) episodes from Seasons 1, 2 and 3 from the American sitcom *Friends*. Three different tests, always following the same design, were constructed. For two of the studies, a 30-item vocabulary test which aimed at evaluating the learning of colloquial and informal expressions in the episodes was used. It was divided in 15 meaning recognition MC questions, in which learners were shown a colloquial expression in English and had to select the best definition out of three possibilities, and 15 meaning recall open-ended questions, which gave participants a contextualized set phrase and they were required to provide its meaning. For a third study, a very similar test was created, although it was shortened to 20 questions (10 MC and 10 open-ended).

Two of the studies followed a pre- / post-test design and participants were asked to watch two episodes of the TV series every week. The first study aimed at evaluating the effectiveness of subtitles on colloquial language learning and 40 learners took part in it. They were randomly assigned to two groups: English audio and Spanish subtitles or English audio and English subtitles. The main problem with this distribution was that several proficiency levels were mixed within the same group, and both groups had participants from A2 to C1 levels according to the CEFR. Participants watched 13 episodes over seven weeks (two per week) of the selected TV series and the pre- and post-tests described before were administered at the beginning of this 7-week period and just after it. Results of the study showed significant differences between the two groups at the end of the viewing process. Those having been exposed to English subtitles outperformed the Spanish subtitles' group (p=.01). Effect size was found to be medium, indicating that 9% of the variance was explained by the treatment effect. Finally, no interaction was observed between the two groups and proficiency level, showing that differences were not dependent on it. However, it should be noted that gains were not computed, even if it was shown that scores on the post-tests were dependent on the pre-test. The author points out that, as there were many proficiency levels in both groups, results are proof that exposure to authentic audiovisual materials with on-screen text is beneficial for learners at all proficiencies and from many backgrounds (see also Frumuselu, De Maeyer, Donche, & Gutiérrez Colon Plana, 2015 for more information on this first study).

As a follow-up of this first study, Frumuselu conducted another experiment to test immediate gains in colloquial language learning. The same pool of learners (*N*=49) were divided in two groups again: interlingual (English sound and Spanish text) or intralingual subtitles (English sound and text). Learners were exposed to the same 13 episodes as in Experiment 1 over a period of seven weeks (twice a week). In this study, tests tapped into learners' short-term and immediate acquisition of colloquial language and idioms. As it has been previously mentioned, this second test was reduced to 10 MC and 10 open-ended questions, following the same format and style as the previous one. Nevertheless, no pre- or post-tests were administered, only immediate post-tests were distributed after watching each episode. Using a linear mixed effects regression, the study revealed a no significant effect of subtitling, indicating that intra or interlingual subtitles did not make a difference on how participants' scores evolved throughout the treatment. These results point to subtle differences between Experiments 1 and 2, since an effect of subtitling mode is observed in long-term learning, whereas it is not seen in immediate learning. One of the most interesting points that the study shows is the pre-test effect on the immediate post-tests, since those students with a higher proficiency also seem to benefit more from the treatment; furthermore, as time goes by, the gap between high and low achievers increases. Thus, the second main conclusion that can be drawn is that proficiency level does play a role in immediate vocabulary acquisition through audio-visual input, as "high proficient students score higher than low proficient students and their test results increase over time" (p. 187).

Finally, in the third study, data collection took longer (three months and a half) than in the previous two. Time and amount of exposure were doubled up to 14 weeks and, consequently, students watched a total of 28 episodes (as opposed to the 13 viewed in Experiments 1 and 2). Another difference lies on the fact that all participants were exposed to bimodal English subtitles. According to the author, this approach was adopted on the basis of the results of Experiment 1, in which bimodal subtitling proved to be more beneficial than standard subtitling. The same test format was used, although it was again extended to 30 TWs and expressions appearing in the 28 episodes. A series of *t*-tests showed very significant results between the beginning and the end of the experiment. Besides, post-test scores varied and were widely spread, whereas pre-tests scores were limited to a rather small spectrum. Although results point to a greater benefit of sustained exposure, it would have been interesting to see whether similar gains were obtained with the interlingual subtitling condition, as shown in Experiment 1, where students exposed to Spanish subtitles had also made progress.

Overall, the dissertation is taken as an initial step in the examination of the acquisition of colloquial language and set phrases through audio-visual input with on-screen text. Results are optimistic towards its use in foreign or SLA contexts and they advocate for a sustained exposure to such input, since larger differences were obtained in the long-term (i.e., at the end of the experiment) rather than in the short-term (i.e., when participants were tested immediately after watching each of the episodes). Moreover, as seen in the last study, the longer the exposure to multimodal input, the greater the learning of colloquial and idiomatic expressions, although it is unknown if this learning would have been greater –or would have followed a different pattern– under distinct experimental conditions. The author ends by making a call for further research on the effects of audio-visual input in low-level learners and for including delayed post-tests to gauge participants' level of retention of those TWs previously learned.

It is worth pointing out, though, that other studies have not found such positive results for bimodal subtitling. To start with, findings from Stewart and Pertusa (2004) are not that clear. Their study lasted two semesters during which two different cohorts of intermediate learners from Spanish conversation classes saw two well-known Spanish movies with Spanish audio and either English subtitles (42 students in total) or Spanish captions (53 participants in total). Each movie was divided into three segments (for logistic reasons); before and after each one, participants were given a vocabulary quiz in Spanish focusing on words that researchers presumed to be unknown to students. No actual verification whether the words were unknown to learners was made in the second semester, as it was proven that most of them were new to those students participating in the first semester. After gains were computed, it was observed that differences between groups did not reach statistical significance. However, results seem to be inconclusive and they cannot be largely generalized as, for instance, data on learners' proficiency level or language background is missing and no information is given about the characteristics of the tests that were designed for the purpose of the study.

Other recent investigations have also looked at more innovative captioning techniques (e.g., keyword captioning). Montero Perez et al. (2015) was one of the first to compare full captioning (a verbatim transcription of the dialogue of the clips) and keyword captioning (which only included those lexical items that were considered essential by independent readers, in this case representing a 15.4% of the full transcript). They also introduced test announcement as a form to operationalize "incidental" or "intentional learning" (telling learners in advance that they would be tested immediately after the experiment was conceived as the intentional learning condition). Using eye-tracker, the authors further analysed how the two subtitling conditions affected learners' attention allocation to the TWs.

51 undergraduate students (aged 18-19) at a Flemish university studying French as a FL were selected for the study. Following the results of a vocabulary test specially designed for this research, their level was determined to be (high) intermediate. Two short clips of about two to six minutes already used in other studies (Montero Perez et al. 2013, 2014) were used again and, six weeks before the experiment, a pre-test following the VKS format was administered to assess previous knowledge of 18 TWs. Due to the use of authentic materials, TW length, frequency of occurrence or concreteness (among others) could not be controlled for, although it was made sure that all TWs appeared as keywords in the non-fully captioned condition. After watching the two video clips once, while being recorded by a Tobii X120 eye-tracker, learners were given a comprehension task to make sure that they were paying attention to the content of the video and four vocabulary tests measuring: form recognition, clip association (i.e., selecting the video in which TWs had appeared), meaning recall, and meaning recognition (not analysed afterwards). There was a non-significant effect for test announcement in the clip association and form recognition tests. On the meaning recall test, though, results showed a low but significant effect for test announcement, revealing that those who were explicitly told about the upcoming test outperformed the others. Regarding vocabulary learning, results showed that the keyword captioning group outperformed the other on the form recognition test. Focusing on gaze behaviour, while eye-fixation times partially predicted word learning in the full-captioned group, this was not true for the keyword captioning condition. Based on these results, the authors concluded that the visual salience through keyword captioning aided form recognition of the TWs although it did not lead to further processing. At the same time, those who were told about the test in advance seemed to explicitly process TW meanings and this resulted in a deeper semantic process which, at its turn, led to small but significant differences between incidental and intentional learning conditions. The authors comment on the eye-fixation data too to conclude that both enhancement techniques (test announcement and visual

support) have the potential to influence learners' attention (and therefore noticing) of the target items and prove the potential of (keyword) captions as a tool to enhance attention and vocabulary learning in the classroom context.

In sum, most studies with intermediate learners did focus on the effectiveness of interlingual and intralingual subtitles, the latter proving to be more beneficial in most learning contexts. Only one study (Montero Perez et al., 2015) explored one alternative technique (keyword captioning) and it resulted in a greater potential for FL learning (in comparison to the other conditions included in such study). Another point of interest is the role that proficiency level plays (operationalized in some studies as the number of words already known in the TL). As highlighted by Peters et al. (2016), the older or more proficient learners are, the greater their lexical gains when exposed to multimodal input (i.e., the Matthew effect in language learning; Stanovich, 1986), probably due to the larger number of free attention resources they have to allocate to information or input coming from different sources. Moreover, long-term vocabulary learning from extended exposure to video viewing has been found to be greater than immediate learning (Frumuselu, 2015).

4.4.3. Beginner learners.

Only one study that we are aware of has looked at which subtitling technique is better for beginner learners, probably due to the difficulty of exposing these participants to subtitled materials. At this stage, learners may have serious trouble following them. d'Ydewalle and Van de Poel (1999) pooled 327 students from Grades 3-6 in a Dutch speaking school and exposed them to a subtitled short movie. The languages involved were Danish or French (which the participants were learning) and Dutch (their L1). The design totalled four experimental conditions: two being reversed conditions (Dutch audio + Danish subtitles and Dutch audio + French subtitles) and two more being standard approaches (French audio + Dutch subtitles and Danish audio + Dutch subtitles), plus a fifth control condition (Dutch audio + Dutch subtitles). In total, there were 20 groups (five conditions and four age groups; one per grade). All students completed vocabulary, morphology and syntax tests, matched for the modality (aural or visual) and the language (French or Danish) they were exposed to during the treatment. Real but limited language acquisition was found, and language typology / distance was seen to be a good predictor of vocabulary acquisition, since more words were picked up when Danish, a closer language to Dutch than French, was present either in the audio or the on-screen text. Nevertheless, no acquisition of morphology or syntax was observed.

Although the amount of learning observed in lower-level learners was not as high as in the more proficient populations, it is important to highlight that simple exposure to media and TV viewing resulted in some acquisition or, in other words, it was not found to be a negative practice for children’s language development.

Table 4.3
Summary of the studies presented in section 4.4. on vocabulary acquisition from different types of subtitles.

Note. The dotted lines mark a difference in proficiency level (in descending order of proficiency: advanced, intermediate, beginner) of the participants in the studies.

Study	Participants	Experimental design	Input	Vocabulary test	Results
Zarei (2009)	Iranian undergraduates	Watching FL videos with bimodal, reversed or standard subtitling	Nine episodes of a TV series (≈30' each)	40 TWs, recognition and recall vocabulary tests	Bimodal subtitling beneficial for vocabulary recognition and recall
Peters et al. (2016)	(Low-)intermediate high-school learners	Watching FL videos with L1 subtitles vs. captions	13-minute clip	39 TWs, spoken form recognition and meaning recall vocabulary tests	Captioning more beneficial for vocabulary acquisition
	Low-intermediate vocational-school learners		One episode of a TV series (20')	18 TWs, form recall and form and meaning recognition vocabulary tests	Captioning better for recalling vocabulary; no difference in recognition
Zarei & Rashvand (2011)	Adult intermediate university learners	Watching FL videos with L1 vs. L2 subtitles; effects of verbatim vs. non-verbatim subtitles	Summarized version of an American comedy (70')	MC vocabulary recognition test	No difference between L1 / L2 subtitles; non-verbatim subtitles more beneficial
				Fill-in-the-blanks production test	L2 subtitles more beneficial; no difference between verbatim and non-verbatim subtitles
Frumuselu (2015)	Second-year English majors, most intermediate	Effects of L1 vs. L2 subtitles on colloquial language learning	13 episodes of a TV series (≈25' each)	30 TWs, MC meaning recognition test and meaning recall open-ended questions	L2 subtitles more beneficial for vocabulary learning
				20 TWs, MC meaning recognition test and meaning recall open-ended questions	No significant main effect of subtitling on immediate gains
		Effects of L2 subtitles on colloquial language learning (no comparison)	28 episodes of a TV series (≈25' each)	30 TWs, MC meaning recognition test and meaning recall open-ended questions	Significant progress from the beginning to the end of the intervention

Study	Participants	Experimental design	Input	Vocabulary test	Results
Stewart & Pertusa (2004)	Conversational classes intermediate learners of Spanish	Watching movies with L1 vs. L2 subtitles	Two full-length movies	Vocabulary quizzes before and after each movie segment	No difference between experimental conditions
Montero Perez et al. (2015)	(High-)intermediate university learners of French	Video viewing with full vs. keyword captioning; comparing test announcement effect	Two short clips (≈3' and 7' each)	18 TWs, four vocabulary tests including form recognition, clip association, meaning recall and meaning recognition	Test announcement improved meaning recall; keyword captioning better for form recognition
d'Ydewalle & Van de Poel (1999)	Grades 3-6 primary school learners	Video viewing with bimodal vs. reversed subtitling	10-minute short movie	Meaning recognition vocabulary test, sentence construction test and morphology test	Limited vocabulary learning; important role of language distance

4.5. Summary of results on vocabulary acquisition through multimodal input

As seen in this chapter, the studies on vocabulary acquisition through multimodal input have mostly investigated three different facets of FL acquisition: whether video viewing, as opposed to no viewing, is beneficial for SLA (e.g., Peters & Webb, 2018), whether captioned video viewing is conducive to more learning than uncaptioned video viewing (e.g., Rodgers, 2013, study 5), or the most efficient audio textual combination for FL vocabulary development (e.g., Frumuselu, 2015).

To the researcher's surprise, only two studies included a group of students who were not exposed to multimodal input (just took the vocabulary tests), and whose performance was then compared to that of students watching videos. This comparison has often been neglected, even if the field would benefit from knowing if video viewing is conducive to higher gains than other practices (before even enquiring about the role of captions or the effectiveness of different subtitling modes). That made clear, both of them (Peters & Webb, 2018; Rodgers, 2013, study 2) found that video viewing was conducive to more vocabulary learning than test taking. However, the two used quite a homogenous population, since the authors selected intermediate undergraduate learners; hence, very little is known about the potential benefits of video viewing for beginner and advanced learners, or in contexts different from Belgium and Japan, where these studies were conducted.

A large number of studies, however, has focused on whether multimodal input ought to be presented with or without textual support. In general terms, they have found that subtitles or captions are beneficial for the acquisition of FL vocabulary in a variety of contexts and with different populations: for example, Winke et al. (2010) with advanced learners, or Koolstra &

Beentjes (1999) with beginners. However, opposite results have also been reported with intermediate (Rodgers, 2013, study 5) and beginner students (Galimberti, 2016). There is only one study (Hsu, 2013) that found that those exposed to uncaptioned video did significantly better than those who saw the videos with textual support but, as we have already mentioned, it followed a very different approach to other studies on the topic. Overall, and also drawing on the results of Montero Perez et al.'s (2013) meta-analysis, it seems that captions have a positive effect for vocabulary acquisition.

In relation to the subtitling mode, L2 subtitles seem to favour the learning of vocabulary (e.g., Frumuselu, 2015; Zarei & Rashvand, 2011) although L1 subtitles have also been found to be effective as well (e.g., d'Ydewalle & Pavakanun, 1995). It is also worth mentioning that the few studies available with alternative techniques like keyword captioning, highlighted keywords or glossed captioning suggest that they can be even more effective than more traditional approaches (i.e., full captioning) (Montero Perez et al., 2015, 2018).

In sum, the issue of whether video viewing (either with or without textual support) promotes vocabulary learning more than other methodological approaches or teaching techniques has not been analysed to a full extent. What is more, no study to date has explored extensive video viewing as a complement to explicit vocabulary instruction. In addition, the majority of the studies reviewed are cross-sectional and exposed learners to very small amounts of multimodal input, with short clips lasting no more than a few minutes (e.g., Winke et al., 2010). Only three (Frumuselu, 2015; Rodgers, 2013; Zarei, 2009) explored the long-term effects of exposure to video viewing and included several episodes of the same TV series to be able to trace participants' vocabulary learning evolution. However, the three tapped into the same language learners' profile (university undergraduates), and hence very little is known about the long-term effects of video viewing in other populations and in different educational contexts (e.g., primary or high school). Furthermore, we do not know much about the retention effects of the vocabulary learned from video viewing exposure or whether the tenets of the DCT hold true as well for information retention. The few studies which checked retention effects and included a delayed post-test (e.g., Baltova, 1999; Nagira, 2011; Peters & Webb, 2018) administered it just a few weeks after the end of the intervention and inconclusive results were found.

Bearing all this in mind, the study on vocabulary proposed in the present dissertation is mainly a response to: (1) the lack of research dealing with extensive video viewing in classroom instruction, (2) the inconclusive results found with populations other than (intermediate) university undergraduates, (3) the lack of longitudinal research on the topic, and (4) the absence of studies tapping into the long-term retention of the vocabulary learned through multimodal input.

CHAPTER FIVE - CONTENT COMPREHENSION OF MULTIMODAL INPUT

After giving account of what research has found in relation to vocabulary acquisition through video viewing, this chapter will explore content comprehension of multimodal input. First of all, a brief summary of what we understand as listening and content comprehension will be provided. It will also link content comprehension and vocabulary knowledge by analysing what research has found regarding how much vocabulary learners should know in order to successfully understand multimodal input. Then, research on the possible effects of video viewing on comprehension will be presented. A difference will be made, as in the previous chapter, between results found at different proficiency levels. However, and this will be different from what was done in Chapter 4, no division according to research design will be made when presenting the studies, since the majority tend to follow the same pattern (i.e., comparing captions vs. no captions).

5.1. Listening and content comprehension

Listening comprehension is a complex process in which the listener takes an acoustic signal and interprets it based on linguistic and non-linguistic knowledge. It is also an automatic process in which language processing must be extremely fast and efficient because there is not much time to reflect upon meaning, as opposed to what happens with a written text. However, listening does not just entail a mere automatic process of decoding aural language; in other words, the listener is expected not only to extract the meaning of a spoken text, but also to be involved in an active process of inferencing (Buck, 2001). Listening comprehension is also individual and personal; that is, there can be as many interpretations of a spoken text as listeners, especially if the text is complex and ambiguous. Moreover, listening comprehension is better characterised as an ongoing process in which listeners are expected to constantly modify their initial hypotheses as they get new information. It is during this process that different types of knowledge come into play: listeners' knowledge of the world, their past experiences, current thoughts, feelings, intentions, personality and intelligence, apart from their already-acquired language knowledge.

In this respect, Buck (2001) points out that the construct of listening comprehension is composed of two distinct abilities: language and strategic competence. The first involves any knowledge about the language the listener brings in, and it can be procedural, declarative, grammatical, discourse, pragmatic or sociolinguistic knowledge. The second comprises the cognitive and metacognitive strategies that help the listener to use their language competence in a real listening situation. It also includes the compensatory strategies used by L2 listeners to overcome their lack of language competence. Based on this dichotomy, he defines the "default listening construct" as the ability to: (1) "process extended samples of realistic spoken language, automatically and in real time", (2) "understand the linguistic information that is unequivocally included in the text" and (3) "make whatever inferences are unambiguously implicated by the content of the passage" (p. 114). These three abilities should be taken into account when designing any listening comprehension test, as they tap into what is unique about listening comprehension.

It is also important to note that, up to now, we have talked about *listening* comprehension, and not *content* comprehension, precisely because listening comprehension was first defined and studied in relation to oral input, not considering situations in which video or visual support were provided. However, Buck (2001) already claims that including visuals is a very realistic replication of real-world listening situations. Besides, he also manifests that visual support is an important variable in listening comprehension, as it clarifies the context in which the spoken message is interpreted, brings along paralinguistic information (e.g., gestures, speakers' facial and body expressions, etc.) and has the "potential to influence or change the listener's interpretation of the speaker's words in a significant way" (p. 48). In this sense, Hoven (1999) divides this paralinguistic information into three different types: *kinesics*, *proxemics* and *prosody*. Kinesics refers to communicative movements that complement or replace the verbal message (for instance, hand expressions); proxemics is related to the physical distance between the speaker and the listener and, finally, all the paralinguistic information conveyed by the speakers' accent, intonation and rhythm would be what she labels as prosody. It is recommended that these three elements are incorporated into the language classroom, and "explicit efforts must be made to provide learners with information on the kinesic aspects of messages and how to interpret and produce them" (p. 76), since they convey a significant part of the aural message.

In the case of multimodal input, we consider it is no longer totally accurate to talk about *listening* comprehension when referring to comprehension of a video, since the information learners get comes from both the spoken text and also from extralinguistic sources. This is why we will refer to this type of comprehension as *content* comprehension. Other researchers also share this opinion: alternative terms such as "viewing, audio-visual, video, multimedia or watching comprehension" (Schroeders, Wilhelm & Bucholtz, 2010, p. 563) may be accurate too, since all capture the idea that the receiver does not only rely on the linguistic information of the audio to understand the message. However, we believe that the term *content comprehension* is more adequate because the other terms (such as *viewing* or *watching* comprehension) seem to refer more to the comprehension achieved from the audio and video input (not necessarily subtitled). Hence, the term *content comprehension* would give better account of all the means by which learners receive information when presented with multimodal input and, consequently, it will be adopted in the present dissertation.

5.1.1. The lexical demands of multimodal input.

For successful comprehension of any text (i.e., written or aural), it is very important that the input learners are exposed to is at their actual level of competence. For instance, it has been shown that learners would need to be able to reach 95% (Laufer, 1989) or, ideally, 98% lexical coverage (Hu & Nation, 2000) in order to understand and learn from reading. In this respect, most research on reading has found that 8,000-9,000 WFs are needed to understand a written text (Nation, 2006). Regarding vocabulary demands of spoken texts, Nation (2006) also showed that 6,000-7,000 WFs are necessary to follow this type of texts. There is very scarce research on the lexical coverage of multimodal input, but some studies by Webb and Rodgers have looked at the vocabulary demands of both television programmes and movies.

Regarding television programmes, Webb and Rodgers (2009b) analysed a large corpus of 88 British and American programmes from different genres and characteristics (news, drama, sitcoms, science-fiction, older programmes, and children's programmes). They claim that, in comparison to reading and listening, the required coverage to understand television programmes is lower, although it differs depending on the type of programme. When no distinctions are made between television genres, then it is seen that a VS of three thousand WFs would suffice to reach 95% coverage, while seven thousand would be enough to reach the ideal 98% coverage, always assuming that learners are familiar with marginal words (e.g., *ah*, *oh*, *mmm*) and proper nouns. These figures are a bit lower than the 95% coverage found in reading and listening, set by Laufer (1989) as the level for 'reasonable' comprehension, because the authors claim that "the vocabulary heard in television programs may be supported by visual images" (p. 339).

In the same study (Webb & Rodgers, 2009b), small between-genre differences were found at the 95% coverage level, although they appeared to be more important at the 98% level, as some of the genres (e.g., children's programmes) could be easily understood with knowledge of the first five thousand WFs whereas 9,000 WFs were needed to understand news and science-fiction programmes. In a consecutive study, Rodgers and Webb (2011) focused on related episodes of TV series because they hypothesised that more recycling of vocabulary may take place and there may be more opportunities for incidental vocabulary learning than in unrelated episodes. After matching several episodes from six drama TV series and comparing them to randomly selected unrelated TV programmes, it was seen that "the vocabulary load of related television programs is likely to be lower" (p. 700). Furthermore, the potential for incidental learning was higher for related than for unrelated episodes (there was a higher degree of repetition of low-frequency vocabulary and the percentage of WFs encountered once or twice was much lower). These low-frequency words encountered in related programmes are supposed to be easier to acquire in this context than in class, as learners count with a certain degree of background knowledge (familiar storyline, characters, etc.) and the vocabulary is also found in a variety of contexts. In conclusion, Webb (2011, p. 131) claims:

> [It] may be more effective to watch television programs with related content and storylines than programs with unrelated content. Watching similar programs is likely to reduce the lexical burden and may also increase background knowledge which may aid comprehension when viewing subsequent episodes with similar content.

Regarding the vocabulary needed to understand FL movies, a type of input that has been widely used as well in research on comprehension and multimodal input, Webb and Rodgers (2009a) analysed the scripts of 318 movies from different genres (action, animated, classic, comedy, crime, drama, horror, romance, science fiction, war, and Western films). Results showed that knowledge of the first 3,000 WFs (plus knowledge of marginal words and proper nouns) were necessary to reach 95% coverage, whereas knowledge of the first six thousand WFs was needed for 98% coverage. Across genres, no great differences were seen at the 95% level, as three thousand WFs were needed in eight of the eleven genres. However, larger differences, varying between 5K (e.g., in horror and drama films) and 10K WFs (e.g., in animated movies), were found at the ideal 98% lexical coverage level.

5.2. Research on content comprehension of multimodal input

Price (1983) was the first study to tap into how content comprehension can be affected by the presence of on-screen text when watching videos. Although not much information is given about the participants' familiarity with the TL or previous experience with formal instruction, a total of approximately 500 students from different backgrounds and speaking more than 20 different mother tongues were recruited. Half of them were exposed to four video excerpts with captions and the remaining half to the very same videos with no textual support. Besides, half of the students in each group watched the videos once whereas the other half watched them twice. The results are presented in a very succinct manner, but it is made clear that learners, regardless of their ethnicity or language background, benefited from captioned video viewing, even with a single viewing. The author states that these results are proof of on-screen text's facilitative effect, not only as a tool to increase comprehension but also as a sociocultural resource to get acquainted with the mainstream culture. She also advocates for the use of TV programmes in school settings because they provide authenticity of situations and language use, which are typically lacking in traditional classroom instruction.

After these promising results, a large body of research has been compiled in the last decades. This will be summarised in the following sections and will be presented according to the proficiency level of the population under investigation (advanced, intermediate or beginner learners). For each population, studies will also be presented according to the results found in the studies (positive, neutral or negative).

5.2.1. Advanced learners.

Some years after Price started studying the role of captioning on content comprehension, Garza (1991) investigated the use of captioned materials in advanced FL learning. The author selected two typologically distant languages (English and Russian) and authentic materials, since they are more challenging for more advanced students, the population of the study.

Two groups of students were selected: 40 advanced university students learning Russian, and 70 high-intermediate / low-advanced learners of English, with nine different mother tongues. In order to cover a wide spectrum of authentic materials, ten short excerpts (five in English and five in Russian) from five different genres were carefully selected and ten MC questions were prepared for each video. Each segment was seen twice, and comprehension questions were given in advance, so as not to test participants' memory skills. In terms of procedure, testing was done individually and, in each of the TLs, half of the participants saw the five excerpts with captions while the others saw them with no on-screen text. Results showed a substantial increase of correct responses when captions were available. It also seems clear that captions are more beneficial in those genres where comprehension is impaired by external factors such as background noise and less clear dialogue (e.g., music videos). Nevertheless, the author reckons that these results are not generalizable to other populations since there is also a high variability in the number of correct responses depending on the TL (English learners obtained higher scores than Russian learners). The author points out that on-screen text may be a very useful tool to bridge the gap between the development of reading comprehension skills and listening comprehension (the latter being always less developed than the former in FL learning); thanks to captions, learners are believed to be able to "assign meaning to previously unintelligible aural entities" (p. 246).

Garza's finding that captioned video viewing enhances comprehension was also supported by Brett (1997), who concluded that multimodal input was more beneficial than bimodal input for this purpose. He also compared the effectiveness of these two modalities of input on vocabulary recall. He selected advanced English majors in their final year of their university degree, who were exposed to six short aural texts (1.5-2 minutes long) under three different versions: audio-only, video-only (these two conditions without subtitles) or interactive CD-ROM. The main difference between the three was that CD-ROM allowed learners to receive immediate feedback and have control over digital video and access to online definitions and subtitles. The aural texts were accompanied by some pre- and while-watching tasks, in which learners had to answer True or False (T/F) questions and, in one case, order sentences chronologically. Inferential statistics showed that multimedia presentation was significantly beneficial for comprehension in three of the six texts chosen. Therefore, the study points to a beneficial effect of instant feedback, as learners could monitor their own learning while watching the videos; besides, higher levels of comprehension were achieved using the CD-ROM and it also facilitated listening skills' development to a higher extent than traditional pen-and-paper tasks.

The results outlined by Garza (1991) and Brett (1997) were also confirmed by a couple of studies already explained in Chapter 4, which also investigated how content comprehension could be affected by the presence or absence of subtitles. Just to remind the reader, Etemadi (2012) explored content comprehension of two English documentaries and prepared a MC comprehension test for each of them. Half of the participants were exposed to audio-visual materials with bimodal subtitles whereas the other half was exposed to the same video input without any kind of textual support. Even though results for vocabulary recognition were not statistically significant, a *p*=.000 significance was observed in the comprehension tests, indicating that bimodal subtitling was more beneficial than the mere presence of video input, as well as a good tool to assist the comprehension of movies and TV programmes. In the second study, Winke et al. (2010) also found that comprehension was better with the presence of captions. It is worth remembering that learners of typologically different languages (Russian, Spanish, Arabic and Chinese) were recruited, and all of them saw the video excerpts twice. The authors elaborated a MC comprehension test about the main points of the stories shown to participants and results suggested that comprehension was better when participants could resort to on-screen text, although watching the videos first with captions and then without (or viceversa) did not affect learners' behaviour (what mattered was the presence of captions in one of the viewings). Furthermore, proficiency influenced students' level of comprehension, as was expected.

It seems, then, that similarly to what happened with vocabulary, captions help in the process of language analysis and decomposition and to parse and segment the aural forms presented in the videos. However, it should be borne in mind that, for the experience to be as beneficial as possible, the materials used should match learners' proficiency level and their ability to cope with complex information.

Not all studies prove that on-screen text is beneficial for content comprehension. Some other investigations found inconclusive results which neither prove nor refute a beneficial effect of the presence (or absence) of captions. Among them, Lavaur and Bairstow (2011) compared the behaviour of participants from three different proficiency levels on the same task and exposed to the same video materials. They selected 90 high-school French participants, varying in age from 15 to 18 years old, with three distinct levels in the TL: 30 of them were beginner learners who were not being taught English language at school, 30 more were being taught English language for 3 hours a week and were considered to be intermediate learners, and another 30 were enrolled in international classes and received 50% of the instruction in English (their level was deemed to be advanced). It is important to mention that general proficiency level was not actually tested, as it was operationalized as participants' 'fluency level' in the TL, calculated with a written translation test (French to English and vice versa) and a self-filled-in questionnaire. A short extract from a 1950s film was selected and three different versions were prepared: the original version, with English audio and without subtitles; an intralingual presentation with English audio and English subtitles; and a version with L1 subtitles. At each level, a third of the participants was randomly allocated to each condition. After watching the extract, learners were asked to complete a 42-question comprehension test (half of the questions referring to the dialogue and the other half to the images seen on the screen), followed by a questionnaire.

Inferential tests showed that advanced learners exposed to the original version condition, with no textual support, significantly outperformed the other two groups, indicating that on-screen text was not especially beneficial for the comprehension of the passage, pointing to a distractive effect of subtitles when learners have the adequate proficiency level to cope with multimodal input. Regarding the results of intermediate high-school students, they did not show a significant main effect for type of subtitles. This lack of main effect, which evidences that watching a subtitled movie is neither beneficial nor detrimental for content comprehension in this sample of learners, could be due to the fact that participants had the sufficient level to follow a movie without subtitles, although their capacity was somehow limited to cope with either subtitles *or* (as opposed to *and*) aural input, making it impossible to attend to both at the same time. Besides, as they were told that there would be a vocabulary task after the viewing, they may have focused on achieving a high score on the test, leaving aside other aspects of video comprehension. Finally, beginner learners seemed to process visual and dialogue questions quite differently. On the one hand, watching the film without textual support led to the highest ratio of correct responses to visual questions, the difference being significant between intra and interlingual subtitling. On the other hand, dialogue questions were best answered when the film was presented with English audio and French subtitles. These results suggest that low-proficient learners rely on visual data in order to understand the input they are exposed to, but if subtitles are presented in their mother tongue, reliance on image decreases. The authors conclude that “visual data are critical for low-proficiency students when attempting to comprehend a movie in another language” (Lavaur & Bairstow, 2011, pp. 460-461). Besides, they put forward the idea that subtitles are the key to an increase in global comprehension by those with a low mastery of the language of the film, since they are able to process the dialogue and understand some of the words thanks to the visual and contextual prompts they are simultaneously exposed to.

As seen in this subsection, not many studies have been conducted with advanced learners. Although vocabulary acquisition was found to improve with exposure to multimedia input in this population, not such conclusive results can be drawn as far as content comprehension is concerned. From the different studies with advanced students reviewed here, it can be said that multimodality seems to be positive for content comprehension, although one article (Lavaur & Bairstow, 2011) found a negative effect as well. Next, we will see what has been found with learners at lower levels.

5.2.2. Intermediate learners.

Some research has suggested that captions / subtitles are beneficial for enhancing content comprehension in intermediate learners. The first studies date back to the end of the 1990s (Baltova, 1999; Huang & Eskey, 1999), which were followed by some studies years later with a very similar design comparing captioned and uncaptioned video viewing (e.g., Hayati & Mohmedi, 2011; Markham, Peter & McCarthy, 2001). Some more (e.g., Guillory, 1998; Montero Perez et al., 2013; Park, 2004) explored an innovative technique such as keyword captioning, although not all of them found a positive effect of textual support for comprehension of multimodal input.

To start with, mirroring the design of Garza (1991) and Brett (1997) with advanced learners, Huang and Eskey (1999) investigated whether Closed-Captioned Television (CCTV) viewing led to more developed listening skills and better general comprehension of the episode watched with or without subtitles. 30 intermediate ESL students enrolled in university summer courses were recruited and watched a non-authentic episode of a series designed for language learners. Participants were equal and randomly assigned to one of the two conditions: experimental (which consisted in watching the closed-captioned episode twice) or control (consisting in watching the episode in a traditional uncaptioned format, also twice). After the second viewing, all participants were asked to take a listening test built on the passage seen and consisting of 16 MC questions with three possible answers. Although the statistics reported in the article could have been clearer (e.g., it is unknown which two groups are being compared at some points of the article), the results show that students exposed to CCTV outperformed the control condition on the listening test. Captions did not only help students to understand the story better, but they also contributed to improve students' listening skills. However, although the authors state that the article provides "strong evidence" (p. 86) that CCTV is beneficial for ESL learners' listening comprehension, further research needs to be conducted with other populations at this level using authentic materials.

This is precisely what Baltova (1999) did, as she asked her participants to watch an authentic documentary describing a scientific expedition to the High Artic and she studied comprehension and retention. The comprehension test was administered twice (immediate and delayed test) irrespective of the condition (traditional, bimodal or reversed) to measure not only learning but also how much students remembered two weeks after the end of the experiment. The test consisted of eight open-ended questions which required to understand from one to four idea units in order to be answered appropriately. Statistical analyses revealed a main effect for condition, indicating that students in the two experimental conditions (bimodal and reversed subtitling) outperformed those in the control / traditional condition, although they also experienced greater attrition, precisely because they had obtained higher scores on the immediate post-test and there was more room for comprehension decay. Nevertheless, it was observed that those groups exposed to subtitled materials "demonstrated significantly higher content maintenance relative to the comparison group" (pp. 82-83). When the two subtitled groups were compared, no differences were observed, and a similar attrition rate was seen. Thus, it may be concluded that "the bimodal and the reversed treatments were equally supportive of students' understanding and retention of the information presented in the video documentary" (p. 122). Moreover, thanks to a background questionnaire, it was observed that prior experience with French TV did not decisively affect content comprehension.

In line with Baltova (1999), Markham et al. (2001) also found that watching a subtitled documentary of a similar length was positive for comprehension; even though they restricted participants' exposure to only one viewing. They examined the comprehension of DVD passages by 169 intermediate university learners of Spanish as a FL in three different conditions: with English subtitles, Spanish subtitles or no text at all. The video that participants watched was a seven-minute extract of a documentary about the Apollo 13 mission. After watching the passage in the corresponding condition, learners were asked to write a summary in a free recall task and to answer a ten-item MC comprehension test. The MC test was easily scored by giving one point per correct answer whereas the free summary was divided into idea units, understood in the article as "any thought or idea related to the passage content" (p. 442). Analyses showed important significant differences in both measures and further comparisons established that the absence of captions significantly hindered the comprehension of the DVD passage as measured by the MC test and the free recall task. Those exposed to Spanish (L2) captions significantly outperformed the control condition, although they fell short when compared to those who watched the DVD passage with English subtitles. In other words, standard subtitling proved to be better than bimodal subtitling for intermediate university students learning Spanish. In light of the results, the authors propose that the 'ideal sequence' would be for learners to watch the same video passage twice: first using L1 subtitles followed by a second viewing with target-language captions. If successful, authors also suggest watching the materials a third time without on-screen text.

Following the same design as Markham et al. (2001), Hayati and Mohmedi (2011) compared the listening comprehension of 90 first- and second-year university students at an Iranian university, who were equal and randomly divided into three groups: English (TL) captions, Persian (L1) subtitles, and no on-screen text. The study lasted a total of six weeks in which six small excerpts from a BBC film were shown to students (one every week) in the corresponding condition. At the end of each viewing session, students were given a MC comprehension test consisting of 10 questions based on the content of the passage. The average of the six comprehension tests was calculated and it was shown that students watching the episode with English captions significantly outperformed the other two groups. At its turn, the Persian group outperformed the no-captions group, confirming the general tendency already outlined by previous research in this section, which indicates that the presence of textual support tends to be beneficial for the comprehension of media materials. The authors emphasize the importance of initial proficiency level since it is argued that participants in the English and Persian groups benefited from the intervention thanks to their developed listening skills, which were good enough to decode the chunks of English language they were exposed to. Besides, since students exposed to L1 subtitles had to additionally translate words from Persian to English to chunk the flow of aural input, this could have required more cognitive resources they may not be available, thus decreasing participants' capacity to make sense of what was being said in the video. According to researchers, this additional process of translation (from L1 to L2) may have hindered understanding, whereas the presentation in English could help participants to match the aural and written forms of words, and this consequently boosted content comprehension.

Very similar results were observed by the two master theses already presented in the chapter on vocabulary (Aurstad, 2013; Kvitnes, 2013). Both studies coincide with the idea that there is no difference between bimodal or standard subtitling; what matters, according to the authors, is the presence of on-screen text. Kvitnes (2013), using the sample of 16-year-old participants, administered a comprehension test immediately after learners had watched the episode in one of the three experimental conditions: English (TL) subtitles, Norwegian (L1) subtitles and no textual support. Comprehension rate was quite high in all three conditions (78-90%), although the group exposed to uncaptioned video materials did poorer than the other two. Norwegian and English subtitles significantly contributed to explain the variation in the results. Furthermore, previous vocabulary knowledge was one of the predictors that accounted for a higher percentage of the variance. In the other master thesis, Aurstad (2013) followed the same design and used the same comprehension test, although learners were one year older. Results showed that experimental condition (captioned vs. uncaptioned) was the second factor explaining a higher percentage of the variance. It was found out that watching the episode with on-screen text led to higher comprehension benefits than watching it without subtitles (Norwegian and English subtitles were predictors of the results at the $p<.01$ level). In brief, the two studies found very similar patterns and showed that watching the episode with L1 or L2 subtitles led to higher comprehension gains, indicating that on-screen text helped learners to decode the aural information presented in the series. As the CG lacked the verbal channel, participants in this last condition could only integrate information from two sources and this, according to the authors, led to lower comprehension. It can be added that, according to the results of the studies, a one-year difference (i.e., more formal instruction, higher proficiency level, etc.) does not affect comprehension of media materials (in this case, one episode of the TV series *Family Guy*).

As pointed out at the beginning of the section, Guillory (1998) conducted a ground-breaking study exploring the keyword captioning technique for the first time. She compared the effectiveness of full, keyword and no captioning on comprehension with second-semester university learners of French (*N*=202). Keyword captioning included a 14% of the whole verbatim transcript, identified by experienced raters as being essential to understand the main idea of the audiovisual material. Two video clips specially designed for language learners as accompaniment of the textbook (i.e., non-authentic materials) were chosen and shown to participants only once. Immediately after watching each video, participants were asked to answer a seven-question comprehension test containing open-ended questions. A one-way ANOVA determined a significant main effect for condition, with the captioning groups outperforming those learners in the CG. However, full captioning did not prove to be more effective than keyword captioning, leading the author to think that learners do not need to be subjective to big amounts of text in order to understand the storyline of audiovisual materials. It is concluded, though, that the degree of comprehension learners achieved (52-72%) is not good enough and the author attributes it to the interference caused by captions in learners' attention capacity and to insufficient information included in keyword captioning.

Park's (2004) doctoral dissertation went beyond Guillory's research since he also tapped into keyword captioning but distinguished between three groups of Korean intermediate learners (*N*=89): below-, pure-, and above-intermediate. Participants watched six short authentic videos: two with full captions, two with partial captions and two more with no on-screen text. However, the percentage of words included in the partial captioning condition is not reported. Listening comprehension was measured with three different tasks administered after watching each video: (1) a gap-filling summary, (2) MC questions, and (3) word recognition tests. In (1), learners were asked to read a summary of the short-clip and provide five missing words, in what resembles a vocabulary recall test rather than a listening comprehension activity. Test (2) consisted of five questions with four possible answers each and, in Test (3), learners were asked to listen to the videos again and fill in some blanks with the exact wording.

As expected, the high-intermediate group significantly outperformed the other two in the three experimental conditions. Interestingly, full captioning was the condition that led to the highest comprehension rates in all groups, regardless of their level. Furthermore, significant differences were found between the full captioned and the uncaptioned groups at all levels, whereas the high-intermediate group was the only one in which the keyword technique was equally beneficial as full captioning, indicating that learners need to attain a certain proficiency level in order to benefit from partial captioning. Finally, differences between partial and uncaptioned conditions were found in the upper- and pure-intermediate level, but not in the lowest one, showing the limits of this practice for content comprehension. These results suggest that the processing of keyword / partial captions is a challenging activity for EFL learners, and they are not recommended unless students are at an advanced level, probably due to the need of better listening skills as less textual support is provided.

Montero Perez et al. (2013) ratified the results obtained by Park (2004) even though they used a very comprehensive test that analysed listening comprehension from different perspectives. In brief, they investigated whether full or keyword captioning, in comparison to no on-screen text, led to better comprehension, as well as students' perceived usefulness of such techniques for FL learning. A large number of undergraduate Dutch learners of French at a Belgian university were asked to participate in the study. They were randomly divided into the three experimental conditions: full captioning with French audio and subtitles (*n*=81); keyword captioning with a 14% of the entire transcript (*n*=75), which was considered essential by five lecturers of French; and no on-screen text (*n*=70). Regarding the videos, three short clips (between 2 minutes 30 seconds and 8 minutes) already piloted in previous studies at the same institution were selected as they were considered appropriate for participants' level. Besides, a listening comprehension test was tailored for each of the three clips, totalling 43 questions. The test was rather extensive and included classification, MC, who-says-what, fill-in-the-blank and open-ended questions. It also included both receptive and productive questions, as well as a variety of global and detailed enquiries.

Comprehension tests' quantitative analysis showed a main effect for condition. Post-hoc comparisons established that the full captioning group outperformed the other two conditions, while no difference was found between keyword and no captioning conditions. The same tendency was observed with global questions whereas detailed questions were answered in a similar way irrespective of the experimental condition. Further analyses also revealed a large effect of previous vocabulary knowledge in the two types of questions. In conclusion, these results confirm the general tendency in research that captions are beneficial to foster learners' listening comprehension. However, the lack of difference between keyword and no captioning groups may be attributable to the clips selected for the study, since they may have included too challenging vocabulary and students may have paid more attention to it, rather than to the general storyline of the clips. Moreover, these short clips might have forced participants to rely on their knowledge of English vocabulary because they did not have enough support (either visual or aural) to process the audio input.

However, in a subsequent study following a very similar research design, Montero Perez et al. (2014) did not find any positive effects of keyword captioning. The authors divided the whole cohort of participants (*N*=133), all of them intermediate undergraduates, into four conditions: no captioning, full captioning, keyword captioning, and full captioning with highlighted keywords. Learners were shown three clips (each of them two times) and their job was to complete three comprehension tests. These included both general and detailed questions, and consisted of three exercises: open-ended and T/F questions and a combination task, in which learners had to match statements with the corresponding pictures. Statistical tests revealed that comprehension did not significantly differ across the four conditions; what is more, the no captioning condition obtained higher scores than some of the other conditions (e.g., keyword captioning). Besides, scores on the three comprehension exercises were positively correlated with learners' VS (the larger one's vocabulary, the better their comprehension). The authors suggest that the clips or the questions may have been too easy for the population under study, since the latter could not tap into the target, more challenging vocabulary (otherwise, the comprehension questions would have hinted at the answers of the vocabulary tests) and inferential questions could not be formulated due to the nature of the videos.

In line with this last piece of research, Rodgers (2013), one of the most comprehensive studies in this field, also failed to find significant differences as far as listening comprehension is concerned when comparing captioned and uncaptioned FL television viewing. His thesis presents two experiments exclusively devoted to the study of listening comprehension: in the first one (Study 1), all participants watched 10 uncaptioned episodes of the same TV series (*Chuck*) whereas, in Study (5), another group was added, and the author established the comparison between captioned and uncaptioned viewing. The same order of the episodes shown in the vocabulary study (see Chapter 4) was also followed here, with the peculiarity that the first and last episode were not consecutive, although the eight episodes in between followed the original broadcasting order. The final sample for this first experiment was of 321 intermediate Japanese EFL learners who had been previously taught English for a minimum of seven years. So as to make sure that content comprehension, rather than memory skills, was being tested, and bearing in mind that episodes were on average 42 minutes long, Rodgers (2013) divided each episode in six shorter segments of seven minutes and designed a comprehension test for each of them.

As this study is of special relevance for our thesis, we describe how comprehension was tested in this case and an accurate description of the results obtained is also given. The researcher included three types of comprehension questions: topic questions that could be easily answered while watching the video (7.5%), detailed questions (50.3%) and questions aimed at measuring participants' inferential ability (28.7%) (the remaining 13.5% were sequencing items). The different tests consisted of two exercises: a varying number of T/F statements and MC questions with three distractors each. Finally, at the end of the last segment of each of the episodes, a third comprehension exercise was administered in which participants had to order twelve sentences chronologically to measure global comprehension of the episode. Regarding this last exercise, sentences summarizing the episode were presented, although the sentences in the first and seventh place were already provided by the researcher (among other reasons, to avoid getting a zero score). Thus, students were asked to order 10 of the 12 original sentences. The complete comprehension test was translated into Japanese to make sure that it was comprehension, and not FL ability, which was being tested.

Regarding Study (1), Rodgers concluded that there was a significant main effect of viewing English-language television from the beginning to the end of the intervention, with students progressing on average an 8% from the initial episode of *Chuck* to the final one. Rodgers also analysed how the eight successive episodes of the TV series varied in terms of comprehension and a large degree of variability was found, indicating that the extent to which participants understood the audiovisual materials was episode-dependent. Moreover, there was a big difference between minimum (21%) and maximum (95%) scores in all episodes, although the average comprehension level was situated between 60-70%. As inferable from this data, participants' comprehension did not confront to a single homogenous pattern; on the contrary, it followed irregular patterns that were episode- and participant-dependent. Rodgers also finds that vocabulary knowledge is significantly correlated with listening comprehension ability, although for nine of the ten episodes this correlation was rather small. The importance of serial episodes for the understanding of FL material is also emphasised since "participants may have been able to make use of gradual accumulation of background information resulting in superior comprehension of the final episode" (p. 54). Moreover, the fact that comprehension scores varied was somehow expected since episodes are never of the same difficulty, as there are many factors that play a role in them: accent, hesitations, prosody, utterance length, interest in the plot, image-audio synchronicity, etc. When analysing the role of previous vocabulary knowledge in the comprehension of media materials, it seems that VS and depth are more decisive in the understanding of short clips, rather than in longer videos, as this lack of vocabulary knowledge is compensated by extralinguistic information such as the names and relationship between characters, previous events in the series that work as background information, or the presence of images to confirm or discard previously-formed hypotheses.

In the same way as the author did with the vocabulary study, he also investigated the effectiveness of captions in the comprehension of TV series. Taking the results from Study (1) as baseline (they watched the series without captions), participants' performance was compared to that of an additional group of 51 students, who were exposed to the same 13 episodes as well as received textual support in the form of English captions. Statistical analyses showed that there was a significant increase in the comprehension of captioned TV series from the first to the tenth episode ($p<.001$), revealing that participants' average scores increased a 3% from the beginning to the end of the viewing period.

In agreement with Study (1), it was clearly seen that episode comprehension with captions differed depending on the episode analysed, with the average varying from 60% to 73%, and maximum scores reaching 94% while minimum scores were set at the 32% level. When comparing the behaviour of participants in the two studies (captioned vs. uncaptioned), gains from the initial to the final episodes were computed and it was seen that the no-captions group outperformed the captions group. This indicated that watching uncaptioned episodes led to larger progress than watching the same episodes with on-screen text. Although these findings contradict previous research (see the articles reviewed earlier in this and previous sections) the author justifies it by saying that learners could have resorted to other information sources to compensate for the lack of L2 knowledge. However, it must be acknowledged that episode scores of the captions group tended to be higher than those of the other group, although *t*-tests only revealed significant differences in three of them, especially in those episodes which turned to be more difficult for learners in both experimental conditions. Furthermore, in contrast to the results of the uncaptioned group, in which vocabulary knowledge was a predictor for almost all episodes, such factor proved to play a significant role in only four episodes when audiovisual input was viewed with captions. Thus, the author concludes that "greater vocabulary knowledge may be more important for comprehension of episodes of television when captions are not present" and that "captions allow learners with less vocabulary knowledge to utilize their knowledge of the written forms of the words for comprehension" (p. 258). The same methodology, results and discussion are also outlined in an article published some years later with the data collected for Rodgers' doctoral dissertation (for more details, see Rodgers & Webb, 2017).

In sum, some of the research summarized in this subsection indicates that captions are effective to boost intermediate learners' content comprehension of audiovisual materials (e.g., Huang & Eskey, 1999). Other studies have not revealed a significant effect of type of captioning or have failed to show the potential of new techniques like partial captioning or highlighted keywords (e.g., Montero Perez et al., 2014; Rodgers, 2013). Besides, more comprehensive studies investigating the evolution of comprehension after watching several episodes of the same TV series found neither positive nor negative results (i.e., captions were beneficial in more difficult episodes, but they did not have an effect in others). Taking Baltova (1999) and Rodgers (2013) as exceptions, research has not paid much attention to the study of how content comprehension develops with sustained exposure to television. This lack of longitudinal research, and the absence of baseline data to which new empirical research can be compared, is what makes the present study especially worth conducting.

As we have seen so far, most of the research has been carried out with adult participants, probably due to the easy access that researchers have to this population. However, there are some studies focusing on comprehension of multimodal input by younger beginner learners. In the section that follows, research on beginner learners and content comprehension of multimodal input will be presented.

5.2.3. Beginner learners.

Research reviewed in this section has not only concentrated on children or adults with an initial level of the TL (e.g., d'Ydewalle & Pavakanun, 1995; Galimberti, 2016), but has also explored other facets of the effects of video viewing on listening comprehension (e.g., the type of comprehension being fostered –Lee, Roskos, & Ewoldsen, 2013– or the development of general listening skills –Becker & Sturm, 2017).

To begin with, d'Ydewalle and Pavakanun (1995) looked at how content comprehension varies when learners are presented with different types of subtitles (in English, Dutch and control condition) in combination with three audio types and in two different populations: adults and adolescents. In the first experiment, conducted with English university learners with a beginner level of the TL (Dutch), participants had to answer a 33-item comprehension test based on a 12-minute clip. After a multivariate analysis of the results was conducted, the language of the soundtrack proved to have a significant main effect in the comprehension test. As expected, greater comprehension was achieved with English soundtrack, followed by the Dutch version and the no-audio condition. A very similar pattern was observed with the presence of on-screen text, as better comprehension was achieved when English was present in the textual support. Overall, it seems that both standard and reversed subtitling conditions led to similar degrees of content comprehension. In Experiment 2, when authors followed the same procedure but selected 99 Grade 9 adolescents, findings indicated that comprehension was significantly affected when subtitles were not available and improved when English was present either in the soundtrack or the subtitles. Dutch presence in the subtitles also enhanced content comprehension. These two experiments show that participants clearly benefited from the presence of subtitles, since the control condition (no subtitles) obtained the poorest performance both with adults and Grade 9 language learners. Finally, it should be noted that, despite tapping into different constructs, scores on the vocabulary and comprehension tests were positively correlated in both experiments, indicating that a good level of comprehension needs to be achieved (see Webb & Rodgers, 2009b) before vocabulary learning from mere exposure to TV series takes place.

Apart from the benefits that subtitles entail for listening or content comprehension, studies have suggested that subtitles especially promote local comprehension of the audiovisual input, as opposed to more global coherence. Lee et al. (2013) explored the impact of subtitled video on the number of inferences generated during its viewing in two experiments. Although the degree of content comprehension derived from watching movies with on-screen text was not quantified, their study constitutes a new line of research on multimodal input. In Experiment 1, 27 true beginner learners of French, who could neither read nor speak the TL, were divided into two conditions: standard, in which the movie was presented in participants' L1, or subtitled condition, when the selected movie was shown with French audio and English subtitles (i.e., standard subtitling). A classical movie from the 1950s was chosen and only the second half (45 minutes) was used, whereas the events that took place in the first half were summarized and given to students. Participants were instructed to stop the movie at any point they had any thoughts about what was going to happen next or what had already happened and to write them down. These inferences were classified depending on whether they referred to prior information, to upcoming events, or elaborated on an immediate scene (i.e., bridging inferences). Significant differences were found in the number of elaborate inferences (more common in the dubbed condition), but not in the number of bridging inferences (present to the same extent in both conditions). The authors argue that "watching the subtitled version produced higher local coherence and lower global coherence than watching the standard [unsubtitled] version, and vice versa" (p. 422). In a second experiment, 53 participants with very similar characteristics to those in Experiment 1 were asked to watch the video and group the 76 events in which the extract had been previously divided by the researcher according to how similar the events were to one another. Models built on the results from this second experiment also showed that uncaptioned video viewing increased global comprehension of the movie whereas subtitled viewing tended to foster local coherence, in what seems a corroboration of the results from Experiment 1.

Overall, findings suggest that the two groups comprehended the movie differently, with a bias towards global comprehension by the dubbed group while a more local / detailed type of understanding was fostered in the subtitled group. According to the researchers, when people watch a FL film, memory and attention resources are taxed and they tend to focus and make sense of events that are temporally close. This would explain why local comprehension was better, at the expense of more general / global coherence.

However, as was the case with more proficient populations, research has also found that subtitles did not actually lead to better comprehension than uncaptioned videos among beginner learners. There are two studies where results are not that clear: one aims at determining which subtitling mode, if any, leads to a higher degree of comprehension (Galimberti, 2016), the other at checking whether exposure to video viewing promotes listening skills (Becker & Sturm, 2017).

In her master thesis with Grade 6 Italian children (A1 / A2 level), Galimberti (2016) explored the effectiveness of standard and bimodal subtitling (in comparison with a non-captioning condition) on learners' content comprehension of an episode of an American TV series targeting children. Partially inspired by Rodgers' thesis, the comprehension test consisted of three exercises: five T/F questions, five MC questions with three possible answers each and an event-ordering exercise in which learners had to put eight sentences in the correct order, although the first and the fifth were already given. The comprehension test was administered in last place, after the vocabulary test, to avoid any interference with the TWs selected from the episode. The mean scores for all three groups were very similar, and inferential statistics ratified the lack of significant differences in the results. The author concludes that this poor benefit of on-screen text may be due to participants' limited proficiency and cognitive capacity. The fact that participants came from a dubbing background (Italy generally dubs all content shown on TV) implied that students were not familiar with this type of input, which made the activity specially overwhelming for them. The author notes that standard deviations for participants' scores on the comprehension test were rather high, especially for the event-ordering exercise although, when this exercise was excluded from the analysis, the same results were obtained.

Finally, Becker and Sturm (2017) tapped into the integration of audiovisual materials in the FL classroom and adopted a repeated-measures design to analyse how listening comprehension can be affected in the short- and medium-term. A total of 31 English beginner university students learning French during their second semester were divided into two groups: the EG, who completed daily activities using audiovisual media for five weeks, and the CG, who practised their listening skills following traditional classroom instruction. In each of the 13 sessions, 10 minutes were devoted to the practise of listening skills, the difference being that, for example, the EG watched some clips of commercials, series, movies on the topic being discussed, while the CG answered orally to questions posed by the researcher, completed sentence tasks or was involved in group discussion. Three testing times were included in the design: a pre-test, on the first day of the learning phase, a post-test on the last day and a delayed post-test one week after the end of the intervention. All tests consisted of 12 or 13 open-ended questions and were administered in the learners' L1 in order to test their listening skills, rather than their ability to write in the TL. One of the weak points of the study is that, to avoid familiarity with testing materials, a different passage was used at each of the three testing times;

thus, results cannot be attributed solely to the treatment but also to the test, as some of the passages may have been easier or more difficult than others. After checking that the two groups' listening skills did not differ at the start of the experimental phase, all scores were converted into percentages and analysed as a whole using a one-way ANOVA. Although the EG obtained higher scores than the CG in the two testing times after the experimental phase, the difference was not large enough to reach statistical significance. Nevertheless, the fact that the EG overcame the initial difference with the CG proves that simple exposure to media materials was beneficial, or at least not negative, for developing students' listening comprehension in a classroom-context.

To conclude, what seems clear from the few studies conducted with beginners is that the presence of the L1 (either in the subtitles or in the soundtrack of the clips) favours content comprehension. In all studies that looked at the comparison between standard, reversed and bimodal subtitling, those groups with L1 support tended to perform better, although differences were not always significant. However, it should be noted that, at this level, fewer positive results are found (in contrast to more advanced learners): there are comparatively more studies finding no differences between watching videos with and without subtitles. This might certainly be due to students' poor reading and listening skills, still in need to be further developed in their L1 (in the case of children) and, undoubtedly, in their L2 and / or to the lack of practice of the students. Interestingly, though, none of the studies summarised here provided beginner learners with large amounts of multimodal input. Hence, it would be interesting to explore whether there is any progression and, if there is, what type of comprehension is being enhanced in beginner learners in extensive viewing conditions.

5.3. Summary of results on content comprehension of multimodal input

Chapter five has expounded and evaluated the research conducted so far on multimodal input and content comprehension (see Table 5.1 for a summary of the studies detailed in the chapter). It is accurate to say that the simple exposure to multimodal input seems to be a good way to foster and enhance participants' content comprehension. Furthermore, advanced learners seem to strongly benefit from subtitles when videos are short or in one-off studies (e.g., Winke et al., 2010); in other words, subtitles appear to be helpful when learners do not have background knowledge about the characters or the setting of the story. As opposed to what happens with advanced students, intermediate leaners seem to almost always benefit from the presence of on-screen text. According to the research outlined in this chapter, learners at this proficiency level still have a limited competence in the TL and need to resort to textual support to fully understand the audiovisual materials they are exposed to (Huang & Eskey, 1999). Nevertheless, when they are given the opportunity to watch a set of serial episodes, then the benefits of captions are not clear, and are just helpful in complex episodes (Rodgers, 2013, Study 5). In extensive viewing, though, content comprehension seems to progress slow, both with and without textual support (Rodgers, 2013, Studies 1 and 5). Moreover, standard and bimodal subtitling seem to be equally beneficial for intermediate learners, showing that the presence of the L1 is not that crucial for this population (Markham et al., 2001), as opposed to what has been found with beginner learners. This latter population seems to only benefit from multimodal input when it includes the L1, either in the textual support or the audio (if subtitles are in the L2) (d'Ydewalle & Pavakanun, 1995). However, it has been shown that, despite the presence of L1 subtitles, beginners' content comprehension of TV materials is relatively limited. It is also true that more innovative techniques such as keyword captioning did not result in better comprehension when compared to full captioning (e.g., Montero Perez et al., 2013, 2014), probably because participants were not proficient enough to understand the rest of the aural text which was not shown on the keyword captioning condition.

Finally, it is important to note as well that one study (Baltova, 1999) tapped into how much of a video participants remembered two weeks after the experiment, and found lower retention rates in the subtitled groups although these had obtained higher scores on the immediate comprehension tests.

It is obvious, then, that TL proficiency plays a determinant role in the comprehension of audiovisual materials, since it limits the extent to which learners benefit from such practice (Park, 2004). However, there is a lack of research connecting content comprehension and extended exposure to audiovisual materials; primarily because the scarce research on this topic has revealed different results when learners are exposed to multimodal input in extensive viewing conditions. Furthermore, studies have typically used excerpts from television programmes or movies, and not much research has been done with unabridged programmes, which are typical of real-life situations. Finally, even though TL proficiency has been shown to affect learners' degree of content comprehension, there are virtually no studies with learners from different proficiency levels following a similar research design. Consequently, longitudinal studies with authentic materials at different proficiency levels would shed more light on the matter and would be very informative to (dis)confirm the results found by previous research.

Table 5.1
Summary of the studies presented in Chapter 5 on content comprehension of multimodal input.

Note. The dotted lines mark a difference in proficiency level (in descending order of proficiency: advanced, intermediate, beginner) of the participants in the studies.

<table>
<tr><th>Study</th><th>Participants</th><th>Experimental design</th><th>Input</th><th>Comprehension test</th><th>Results</th></tr>
<tr><td>Price (1983)</td><td>Around 500 students from different backgrounds</td><td rowspan="2">Captioned vs. uncaptioned video viewing</td><td>Four video excerpts (seen once or twice)</td><td>Information not available</td><td rowspan="2">Captioned video viewing beneficial for video comprehension</td></tr>
<tr><td>Garza (1991)</td><td>Advanced university learners of Russian and (low-)advanced learners of English</td><td>Ten short video excerpts (2’-4’ each)</td><td>Ten MC questions for each excerpt</td></tr>
<tr><td>Brett (1997)</td><td>Final year university students</td><td>Exposure to audio-only, video-only or interactive CD-ROM</td><td>Six short aural texts (1’-2’ each)</td><td>T/F questions and event-ordering exercise</td><td>Interactive CD-ROM led to better comprehension</td></tr>
<tr><td>Etemadi (2012)</td><td>Iranian English translation majors</td><td>Viewing FL videos with vs. without bimodal subtitles</td><td>Two documentaries (20’ and 30’ each)</td><td rowspan="2">MC comprehension test</td><td>Bimodal subtitling beneficial for video comprehension</td></tr>
<tr><td>Winke et al. (2010)</td><td>2nd and 4th year university learners of Spanish, Russian, Arabic and Chinese</td><td>Viewing captioned vs. uncaptioned FL video</td><td>Three short clips (3’-5’ each)</td><td>Captioned video viewing beneficial for video comprehension</td></tr>
<tr><td rowspan="3">Lavaur & Bairstow (2011)</td><td>Advanced high-schoolers at an international school</td><td rowspan="3">FL video viewing with bimodal, standard or no subtitling</td><td rowspan="3">Short extract from a film (≈9’)</td><td rowspan="3">Comprehension test with half of the questions referring to the dialogue and half to the images</td><td>No-subtitling group significantly outperformed the others</td></tr>
<tr><td>Intermediate French high-school students</td><td>No main effect for subtitling</td></tr>
<tr><td>High-school true beginners</td><td>Bimodal and standard subt. beneficial for questions on visual content; standard subt. for dialogue questions</td></tr>
<tr><td>Huang & Eskey (1999)</td><td>ESL students enrolled in university summer courses</td><td>Captioned vs. uncaptioned video viewing</td><td>Non-authentic episode of a TV series (21’) (seen twice)</td><td>MC listening test with three options</td><td>Captioning led to better scores</td></tr>
</table>

<table>
<tr><th>Study</th><th>Participants</th><th>Experimental design</th><th>Input</th><th>Comprehension test</th><th>Results</th></tr>
<tr><td>Baltova (1999)</td><td>Grade 11 high-school learners</td><td rowspan="5">FL video viewing with L1, L2 or no subtitles</td><td>7-minute video (seen thrice)</td><td>Eight open-ended questions (taken twice)</td><td>Subtitling beneficial for comprehension; greater attrition by subtitled conditions</td></tr>
<tr><td>Markham et al. (2001)</td><td>Intermediate learners of Spanish</td><td>7-minute documentary</td><td>Free written summary and MC comprehension test</td><td>L1 subtitling beneficial for video comprehension</td></tr>
<tr><td>Hayati & Mohmedi (2011)</td><td>First- and second-year Iranian university learners</td><td>Six short BBC excerpts (≈5’ each)</td><td>MC test consisting of 10 questions</td><td>Subtitling (esp. bimodal) beneficial for video comprehension</td></tr>
<tr><td>Kvitnes (2013)</td><td>Intermediate high-school learners (aged 16)</td><td rowspan="2">One episode of an animated cartoon (20’)</td><td rowspan="2">18 unfinished sentences with some possible continuations</td><td rowspan="2">Subtitling significantly predicted variance</td></tr>
<tr><td>Aurstad (2013)</td><td>Intermediate high-school learners (aged 17)</td></tr>
<tr><td>Guillory (1998)</td><td>Second-semester university learners of French</td><td rowspan="3">Video viewing with full, keyword or no captioning</td><td>Two non-authentic short videos</td><td>Test with seven open-ended questions</td><td>Captioning beneficial for video comprehension; no difference between full and keyword captioning</td></tr>
<tr><td>Park (2004)</td><td>Below-, pure-, and above-intermediate Korean EFL learners</td><td>Six short videos</td><td>Gap-filling summary, MC comprehension and word recognition tests</td><td>High-intermediate group outperformed the others; full captioning led to better comprehension</td></tr>
<tr><td>Montero Perez et al. (2013)</td><td>Undergraduate Dutch learners of French</td><td>Three short clips (2’-8’ each)</td><td>Different exercises with receptive, productive, global and detailed questions</td><td>Full captioning led to better comprehension</td></tr>
<tr><td>Montero Perez et al. (2014)</td><td>(High-)intermediate Dutch undergraduates</td><td>FL video viewing with full, keyword, highlighted keyword or no captioning</td><td>Three short clips (2’-4’ each)</td><td>Open-ended, T/F and a combination task with general and detailed questions</td><td>No difference between experimental conditions</td></tr>
</table>

Study	Participants	Experimental design	Input	Comprehension test	Results
Rodgers (2013)	(Pre-)intermediate Japanese undergraduates	Viewing uncaptioned FL TV series (no comparison)	10 episodes of an American TV series (≈ 43' each)	Event-ordering exercise and MC and T/F exercises with topic, detailed and inferential questions	Significant main effect of TV viewing; participant- and episode-dependent scores
		Viewing captioned vs. uncaptioned FL TV series			No difference between experimental conditions
d'Ydewalle & Pavakanun (1995)	Beginner university learners	Viewing video with audio and subtitles in the L1, L2, or no soundtrack or subtitles	12-minute excerpt from an animated cartoon	MC comprehension test with 33 questions	Main effect for soundtrack language; subtitling led to better comprehension
	Grade 9 learners in an immersion school				Lower comprehension without subtitles
Lee et al. (2013)	True beginner learners of French	Viewing captioned vs. dubbed videos	Extract from a movie (45')	Writing down the number of inferences from the movie	Dubbing led to more elaborate inferences
				Event grouping according to how similar they were	Dubbing fostered global comprehension; subtitling fostered local comprehension
Galimberti (2016)	Grade 6 Italian learners	FL video viewing with L1, L2 or no subtitles	One episode of a TV series (22')	T/F, MC questions and event-ordering exercise	No difference between experimental conditions
Becker & Sturm (2017)	Second-semester beginner university learners	Practising listening skills with audiovisual materials vs. traditional classroom instruction	Extracts from commercials, TV series, movies, etc.	Open-ended questions (taken three times)	

CHAPTER SIX - RESEARCH QUESTIONS AND METHOD

6.1. Introduction: the SUBTiLL project

The present study has been conducted in the framework of the SUBTiLL project. Very few large-scale projects have investigated the role that subtitling plays in the acquisition of a second or FL (for instance, "Subtitles and language learning", an inter-university project coordinated by the University of Turku and funded by the European Commission). The GRAL Research Group (Catalan acronym for Language Acquisition Research Group) at the Universitat de Barcelona has been carrying out the SUBTiLL project since 2014, whose main objective is to evaluate the effects of watching subtitled videos on EFL learning. The project has explored so far the possible benefits of this practice on vocabulary acquisition (both single words and lexical phrases), degree of content comprehension, segmentation and chunking abilities, and pronunciation of the TL. The SUBTiLL project has been conducted in two phases (2014-2016 and 2017-2019) and has covered a wide range of learners' ages and proficiencies, from primary school children to advanced English majors, and different learning conditions. For instance, it has studied multimodal input and language learning in different environments, from at-home informal settings to lab conditions, although special attention has been paid to the use of FL videos in language classes. It has also explored intentional and incidental vocabulary learning as a by-product of watching English-language television. Besides, many learner-related variables have also been explored in relation to extensive viewing (e.g., age, language aptitude, phonological inhibition capacity or WM) as well as input-related variables such as genre, language of subtitles or textual enhancement techniques, among others. The project, which received funding from the Spanish Ministry of Economy, Industry and Competitivity on two occasions (references: FFI2013-47616-P and FFI2016-80564-R), has also run a large-scale study in the Catalan learning context evaluating the quantity and quality of out-of-school exposure to the English language, more propitious for higher levels of incidental vocabulary learning (Nation, 2013; Webb & Nation, 2017).

The longitudinal study in this thesis, together with Pujadas (forthcoming), laid the groundwork for more research on the topic. For example, Suárez and Gesa (2019) connected the results of the thesis to participants' language aptitude. The present dissertation was part of the main study in the first phase of the project and aims at comparing the effects of video viewing, along with formal curricular instruction, to teacher-led instruction on EFL learning at several proficiency levels. As a natural follow-up, Pujadas (forthcoming) explored which combination of audio and subtitles (standard vs. bimodal) and which learning conditions (focused vs. non-focused) were conducive to greater vocabulary learning and higher comprehension rates.

After the presentation of the main objectives of the SUBTiLL project, attention will now be turned to the RQs that the study seeks to give an answer to and the methodology of the dissertation.

6.2. Research questions

The RQs are presented according to the constructs analysed in the present dissertation (vocabulary learning and retention, and content comprehension) and the three proficiency levels which were inspected. The RQs we aim at answering are the following:

In beginner, intermediate and upper-intermediate EFL learners:

1. Does extensive viewing of subtitled TV series enhance L2 vocabulary learning?
    1.1. To what extent does sustained exposure to subtitled TV series affect vocabulary retention?
2. Does content comprehension of subtitled TV series change across successive episodes viewed?
3. How are L2 vocabulary learning and content comprehension affected by learners' proficiency level?

6.3. Design

The dissertation adopted a between-groups pre-test / post-test design following a pedagogical intervention in two different periods: one academic term (September-December 2015) at university and one academic year (September 2015 - June 2016) in primary and secondary school. The intervention was divided in terms (September-December, January-March, April-June), in accordance with the timing requirements at school (i.e., school holidays at Christmas and Easter interrupted the treatment). Hence, although it was conceived as a year-long experiment, the same procedure was repeated thrice. In each of the three terms, there was a pre-test at the beginning and a post-test at the end. In between, there were seven or eight viewing sessions (depending on time constraints and school calendar). Although the initial plan was to include eight sessions in each term, it was not possible to view all the episodes in two of the terms at high school. In these cases, the last episode was dropped out and the TWs specific to it were not included in the assessment and analysis. A summary of the design can be seen in Figure 6.1.

Level	**Term 1** (Sep. - Dec.)	**Term 2** (Jan. - Mar.)	**Term 3** (Apr. - Jun.)
Grade 6	8 viewing sessions	8 viewing sessions	8 viewing sessions
Grade 10	8 viewing sessions	7 viewing sessions	7 viewing sessions
University	8 viewing sessions		

Pre-test	Post-test T1 Pre-test T2	Post-test T2 Pre-test T3	Post-test T3 ... Delayed test

Figure 6.1 - Outline of the design of the study.

In each of these viewing sessions, several tasks were carried out by the participants in the study, who had been previously divided into two conditions: experimental and control. All students did a vocabulary pre-task in which a series of TWs were introduced. This was followed by the viewing of the episode and a while-watching task. These last two tasks were only performed by the EG. The two groups (EG and CG) did a vocabulary post-task at the end of each viewing session and only those in the EG took a comprehension test based on the episode they had just seen (for more information, see section 6.6.2.). Table 6.1 shows the activities participants performed in each viewing session:

Table 6.1
Tasks performed in each viewing session.

		Vocabulary pre-task	Episode viewing	While-watching task	Vocabulary post-task	Comprehension test
Group	Experimental	✓	✓	✓	✓	✓
	Control	✓	✗	✗	✓	✗

After the presentation of the design of the study, participants' characteristics from all groups will be introduced, with special attention to their language proficiency as well as to other specific requirements to be part of the final sample for the study.

6.4. Participants

The study covers a wide range of language proficiencies: beginner (Grade 6, last year of primary education), intermediate (Grade 10, last year of secondary education) and upper-intermediate (first-year university undergraduates). At all levels, there were two intact classes, each of which was allocated to either the experimental or control condition. With the selection of participants from different grades, we inevitably had a confound between age and proficiency level, since participants at the beginner level were younger than those at the intermediate level, and these, in turn, were less proficient than older students at university. Nevertheless, this confound does not invalidate the study, as it is what is usually found in real language learning environments, in which learners tend to become more proficient as they get older.

The initial sample of participants was of 181 EFL learners, out of whom 102 were allocated to the EGs while 79 formed the CGs. If divided by proficiency groups, there was a total of 52 participants in Grade 6 (27 allocated to the EG and 25 to the CG), 64 in high school (33 belonged to the EG and 31 to the CG) and 65 enrolled at the Universitat de Barcelona (42 in the EG and 23 in the CG). However, strict inclusion criteria were adopted, and the number of participants included in the final sample varied depending on the term analysed. It should be noted that one participant who was considered valid in Term 1 could be excluded from the analysis in Terms 2 or 3, and vice versa, due to the reasons that are presented below.

First of all, all students with language learning difficulties (e.g., severe dyslexia or ADHD), following individual care programmes, or enrolled on an Erasmus programme were excluded from the analyses at all testing times. Additionally, in order to be part of the final sample of participants, it was necessary that:

1) Students had completed the pre- and the post-test for a given term.
2) Students had attended all the sessions of a given term (*n*) or missed a maximum of one (*n*-1).

Based on these criteria, Table 6.2 shows that the final sample of participants comprised 158 learners (90 in the EG and 68 in the CG), divided into three proficiency groups: Grade 6 (40 students - 22 EG vs. 18 CG), Grade 10 (57 participants - 30 EG vs. 27 CG) and university (61 EFL learners - 38 EG vs. 23 CG). The following sections focus, in turn, to each proficiency group.

Table 6.2
Number of participants in the study.

	Group	Term 1	Term 2	Term 3	Delayed test
Grade 6	**EG**	22	23	22	19
	CG	18	17	18	15
	Total	40	40	40	34
Grade 10	**EG**	30	29	28	24
	CG	27	23	20	16
	Total	57	52	48	40
University	**EG**	38			
	CG	23			
	Total	61			

6.4.1. Grade 6.

The initial sample of Grade 6 participants consisted of 52 students throughout the three terms. They belonged to two intact classes, one became the EG (n=27) and the other the CG (n=25). As mentioned above, some participants were excluded from the final sample. Four participants (one in the EG and three in the CG) were excluded from the whole intervention as they presented severe language learning difficulties, hence the initial sample was already reduced to 48 students. After that, as the intervention was divided in three academic terms, the two inclusion criteria were applied per term, independently of whether participants had been included in the previous term analysis or not, that is why the sample sizes vary. Regarding inclusion criterion #1 (i.e., data available for both the pre- and the post-test of a given term), six students (2+4)[8] had to be excluded from the analysis in Term 1, four students (2+2) in Term 2 and 2 participants (1+1) in Term 3. In relation to inclusion criterion #2 (i.e., the number of sessions attended by the participants), two participants (2+0) were excluded in Term 1, four (1+3) in Term 2 and three (1+2) in Term 3. Regarding the delayed test, used to answer RQ 1.1., from the 40 participants (22+18) who were included in the analysis in Term 3, 34 completed this test eight months after the intervention. Out of these, 19 belonged to the EG and 15 to the CG. The remaining six were either not present the day the vocabulary delayed test was administered or had decided to study secondary education in another school.

It should be noted, though, that 19 participants in the EG (73.08%) and 15 in the CG (68.18%) were considered to be valid in all three terms, and these figures were a bit lower when all

[8] The first figure always refers to the number of participants in the EG while the second figure describes the number of participants in the CG.

viewing sessions were considered: 13 participants in the EG (50%) and 5 in the CG (22.72%) had gains computed in all terms and data collected for the 24 viewing sessions.

The three proficiency tests administered to Grade 6 participants (dictation, listening and VS tests; see 'Instruments' section below) showed that the two groups (EG and CG) were at the same proficiency level, since no significant differences were found between them (see Chapter 7). On the dictation test, the EG obtained a mean score of 24 words and the CG of 23 words (out of 50). On the listening test, out of 17 questions, the EG got an average of 12 questions right whereas the average for the CG was 11 correct responses. The VS test (Meara, 2005) revealed that learners in the EG had a mean VS of 1,750 words, while the CG counted with 1,355 words, out of a maximum of 5,000.

Participants in this group were attending their last year of primary education (Grade 6) at a semi-private school in Mollet del Vallès, in Barcelona's metropolitan area. At this level, all learners had had at least 900 hours of formal curricular instruction prior to the beginning of the 2015-2016 academic year. Along the school years, they receive three hours of English per week, with two devoted to formal teacher-led instruction and an extra one, called 'hora B' [hour B], centred upon listening and speaking skills. It was during this 'extra' hour that the pedagogical intervention took place. Participants in this age group were receiving formal curricular instruction in English since Grade 1 (age six), when the teaching of a FL starts being compulsory in the Catalan educational system. 42.5% of participants were males whereas 57.5% were females. In terms of age, 89.1% of participants were eleven at the time of the intervention, while 6.5% of participants were ten and 4.3% were 12, since they were repeaters.

Regarding participants' characteristics that can be relevant to the study, these can be described thanks to a background questionnaire (see Appendix A.1.) that was completed by learners' parents and only one was not returned to the teacher, so the answer rate was very high (97.92%). Students' experience with TV viewing in English was rather low, as only 4.33% of students watched movies or television programmes every week (with a preference for L1 and non-subtitled materials and a dislike for L2 subtitles). Besides, playing computer or videogames in English was a more popular activity among Grade 6 participants, since 32.6% of the learners played with them on a weekly basis, with or without L1 / L2 subtitles. This percentage can be attributed to those gamers who really like this activity and they prefer playing a given computer game, even if they do not fully understand its content and storyline. Listening to music in English was the preferred activity for Grade 6 learners, with 84.8% of them doing so on a weekly basis. In contrast, only 8.6% read books, magazines or comic strips in English. Regarding production skills in English, only 9.78% of participants spoke or wrote in English once a week in different contexts and with different people. However, it is somehow surprising that 26% reported they spoke in English with their friends although this could probably be a sort of game that they play with children their own age. The Internet and the World Wide Web did not foster the contact with the TL: the only two activities that were done online were watching YouTube videos (32.5% doing it once a week) and listening to songs or podcasts (54.4%).

Surprisingly, 82.6% of participants were attending private English classes at the time of the intervention, and 67.4% acknowledged having been attending them for three or more years (first enrolling at the ages of seven or eight). The majority of learners enrolled in private classes were additionally exposed to 1.5 or 3 hours of English per week. Eventually, stays abroad in English-speaking countries and summer camps in English were not common practices among

Grade 6 learners: only 10.7% mentioned having participated in them, with an average length of 15 (stays abroad) or 30 days (summer camps) although this did not guarantee the practice of their speaking skills, since not much contact with English was reported.

6.4.2. Grade 10.

The initial sample of Grade 10 participants consisted of 64 EFL learners (EG=33 and CG=31), from two different classes in the same institution. During the sessions when the intervention was planned, classes were divided in two following learners' alphabetical list ordered by their first surname. For the whole intervention, the first half was allocated to the CG and the second half to the EG. It is important to note that this division was established in order to facilitate the logistics of the intervention and that the groups were always respected; that is, no students were moved from the first to the second half of the class (or vice versa) for research purposes. The inclusion criteria were applied each term, independently of whether or not one participant had been included in the analysis of the previous term.

One student from the CG was excluded from the whole intervention as she presented severe language learning difficulties and followed an individual care programme. In the light of inclusion criteria #1 (i.e., data for both the pre- and post-test of a given term needed to be available), four students (3+1) were excluded in Term 1, eight learners (3+5) in Term 2 and nine (4+5) more in the last term. Besides, some participants attended less than *n-1* sessions and thus did not meet inclusion criterion #2. In this regard, two students (0+2) were additionally excluded in Term 1, three (1+2) in Term 2 and five (1+4) in Term 3. Regarding the delayed vocabulary test, out of the 48 students (28+20) who were included in the analysis in Term 3, 40 participants also took the delayed test; 24 belonged to the EG and 16 to the CG. This implies that there were eight students who did not take the test, either because they were not present the day of the data collection or they had changed schools since they decided to study Vocational and Educational Training programmes, instead of 'Batxillerat' [A levels].

It must be pointed out that 24 students in the EG (72.72%) and 14 in the CG (46.66%) could be analysed in all three terms. Besides, 14 students in the EG (42.42%) and 5 in the CG (16.66%) took all the pre- and post-tests on which the intervention was based and also attended the 22 viewing sessions in which it was divided.

In terms of proficiency, the EG obtained a mean score of 68.09 points (out of 100) on the listening part of the Oxford Placement Test (OPT) (Allan, 2004), whereas the CG's average score was 68.07 points. It is important to note that participants at this proficiency level did not take the grammar part of the OPT. Focusing on the VS tests, they revealed that the EG knew 3,498 words on average whereas the CG knew 3,466 (out of a maximum of 10,000 words). Based on this data, there were no significant differences between the two conditions in terms of proficiency, so the EG and the CG were comparable.

Grade 10 students were attending their last year of compulsory education in the Catalan school system. After successfully completing of this last course, students are given the chance of either studying for two more years before entering university or completing Vocational and Education Training programmes, lasting between one or two years. All participants were enrolled in a secondary school in Mataró, a city within the boundaries of Barcelona's metropolitan area. At this level, participants had received at least a total of 1,100 hours of formal curricular instruction prior to the beginning of Grade 10. During the academic year when the intervention took place, they received three hours of English instruction per week and one out of these three hours was aimed at developing students' speaking and listening skills in a more informal classroom environment. It was during this latter weekly session that the intervention took place. In terms of age and sex, participants were quite homogenous: 91.5% were 15 years old at the beginning of the school year, 3.4% were 14, another 3.4% were 16 as they had repeated one course and 1.7% were 17, since this participant had repeated two courses in his academic life. There was a balanced number of boys and girls: 47.5% were males and 52.5% were females.

Other relevant participants' traits can be described thanks to a background questionnaire (which was different from that used in Grade 6) (see Appendix A.2.). It was completed by learners themselves, to elicit personal information on their knowledge of English and their out-of-class contact with this TL. The answer rate to this background questionnaire was very high (95.16%). On average, 14.66% of learners are familiar with English language TV viewing and they do this activity on a weekly basis. From the different subtitling options available, 17% reported that they preferred Catalan or Spanish subtitles, followed by English subtitles (15.3%) and no textual support (11.7%). Playing English computer games was one of the preferred activities of Grade 10 learners, as 25.5% of them did that at least once a week, and some of them did so on a daily basis. As expected, listening to English music was by far the most popular activity among Grade 10 participants, since a 93.2% reported doing that every week. Reading in English was not such a familiar activity and only 27.2% of respondents claimed to do that at least once a week, although they also acknowledged that it was an activity which was done mostly online (52.5%). There was a high percentage of respondents who claimed that they spoke in English quite regularly, especially with friends (20.3%) or when abroad (33.9%). The Internet was also a popular platform to watch videos or listen to music in English. Moreover, 20.3% of participants answered that they had participated in stays abroad and summer camps where English was involved. Most of these experiences lasted between two and four weeks, during which they had more contact with the TL and could practice their speaking skills. Finally, 57.6% of the students answered that they were or had been attending private classes, some of them for just one year while others had been doing so for many years: students attending extracurricular classes did so during their high-school years and for 2 or 3 hours a week.

6.4.3. University.

The initial sample consisted of 65 freshman students from two different classes: 42 in one group, who were allocated to the experimental condition, and 23 in the other group, who were placed in the CG. Intact classes were respected, even if the number of participants in each group varied, as students could not attend classes in groups they were not registered in. After checking the participants' attendance to the viewing sessions and the completion of pre- and post-tests, it was noted that only three participants (all of them from the EG) did not meet inclusion criterion #1. Besides, another participant from the EG was excluded from the analyses due to the suspicion that she had copied the answers on the post-test. After excluding these participants from the EG, the final sample consisted of 61 students. From these, only three students (1+2) missed one of the viewing sessions, but they were not eliminated from the analyses because they still met the second inclusion criterion.

Regarding the proficiency level of the university group, participants allocated to the EG obtained a mean total score of 133.70 points (out of 200) on the OPT (Allan, 2004): with a mean of 68.70 points (out of 100) in the listening test and 65 in the grammar test (out of 100 as well). Therefore, they were at the B2 level according to the CEFR. The CG obtained a mean total score of 135 points on the OPT (70.35 points on the listening test and 64.65 points on the grammar part). This total score also corresponds to the B2 level according to the CEFR. Regarding VS, measured by the X_Lex (Meara, 2005) and Y_Lex (Meara & Miralpeix, 2006) tests, the EG obtained a mean score of 4,431 words and the CG had a mean VS of 4,269 words. No significant differences were found in any of the proficiency tests taken (see Chapter 7), indicating that the two groups were comparable.

Participants from this group were first-year university students enrolled in Media Studies at the Universitat de Barcelona. All of them were taking 'Speaking and Writing in English' ('Expressió Oral i Escrita en Anglès'). This is a compulsory course worth 6 ECTS credits, it is taught during the first semester of the academic year and, due to the high number of students, they are divided into two groups taught by the same teacher. Following the Catalan educational system, all participants enrolling in a university degree should have at least a B1 level in English, and they should have received a minimum of 1,300 hours of curricular formal instruction. Most participants were between 18 and 20 years old at the time of the experiment. To be more precise, 43.4% of university students were 18, whereas 18.3% were 19 and 15% were 20, with the oldest one being 26 at the beginning of the intervention. This age pattern did not vary across conditions, being 18- and 19-year olds the largest in number in both groups. In terms of gender, females represented 65% of the sample while 35% were male students.

Regarding learners' characteristics, 60 (out of 61) participants (98.36%) completed the online background questionnaire administered at the beginning of the intervention (see Appendix A.2.). In general terms, participants were used to watching English language movies and television series, with 38.3% of them reporting that they watched audiovisual materials on a weekly or daily basis. However, there was a tendency to watch them subtitled in the participants' L1s (53.4% claiming they did so every week), as opposed to L2 subtitles (35%)

and no textual support (26.6%). In addition, participants were very much exposed to English music (88.3% claimed that they listened to it every day), and to some extent to books, newspapers or comic strips in English (49.4% reported reading some of them at least once a week). University participants did not report practicing their English productive skills (namely, speaking) as they had almost no contact with NSs, foreigners or friends with whom they could speak in the TL. It is also worth saying that many of the activities in which English was involved took place in an online environment, mostly the Internet, but also through smartphone apps and gaming sites. Besides, not many students had participated in stays abroad (30%) or English summer camps (16.7%), most of them being between two and four weeks long. Finally, 70% of students claimed that they were taking private classes although not many students had taken these classes for a long time. Most private classes lasted three hours a week on average, although some (13.3%) were longer than this.

6.5. Instruments

Different instruments were used in the present study. Some were specially devised for this pedagogical intervention, while others were standard proficiency tests. To start with, the four TV series used throughout the intervention will be described, followed by the vocabulary and comprehension tests and other tasks designed in relation to the TV series chosen. Finally, the battery of proficiency tests administered will be presented in detail.

6.5.1. TV series.

The TV series had to be proficiency- and content-appropriate for the learners the intervention was designed for. As the study covered three proficiency and age ranges, the same TV series could not be used for all participants and different series were needed. That is why, throughout the pedagogical intervention, a total of four TV series were used. At university level, the study lasted one term so these students watched only one TV series. In Grade 10, the same TV series shown at university was also used during the first term, as the two populations did not differ much in terms of age, preferences and listening proficiency. Then, in Terms 2 and 3, they watched another TV series at the same difficulty level. Participants in Grade 6 were low-proficient learners at primary school and two TV series at their level and more adequate for their ages were selected: the first was watched in Terms 1 and 2, and the second in Term 3.

A summary of the selection criteria applied when choosing the TV series best suited for each level is presented below:

- The TV series had to be easy enough for students to follow although they should also be a bit beyond learners' competence level, what Krashen (1985) defined as the *i+1*. In other words, the TV series needed to have at least a small range of vocabulary presumably unknown to students.
- Episode length was also a point to be considered. As classes lasted between 50 and 60 minutes, and different tasks (including the viewing of the episode) had to be carried out, episodes could not last more than 20-25 minutes. This restricted the choice of TV series and ruled out some genres (e.g., thrillers or dramas) since they are usually between 40 minutes and one hour long. As sitcoms are shorter and do not typically exceed 30 minutes, they were considered the most adequate genre.
- There were also many technical requirements that the TV series needed to meet. First of all, the episodes needed to be available in DVD (not, for instance, in online platforms like Netflix, HBO or Amazon Prime Video). For ethical and legal reasons, they could

not be downloaded from the Internet or copied from any other source. Besides, the DVD had to be edited in region 2 (working in Europe, Middle East, Egypt and South Africa) and should have Spanish (for Grade 6 learners) or English (for Grade 10 and university) subtitles available.

- After scrutinizing the different TV series that met the above criteria, only those that were not being broadcasted in Spain at the time of the intervention were considered. Besides, TV series that had been recently broadcasted and were quite popular among children and teenagers were discarded too. Less popular TV series were deemed to be more appropriate because, if learners were familiar with the series chosen, the scores on comprehension tests at the end of each session may have been affected. Therefore, bearing in mind all these factors, the following TV series in Table 6.3 were selected:

Table 6.3
TV series selected for the pedagogical intervention.

<table>
<tr><th>Level</th><th>Term 1
(Sep. - Dec.)</th><th>Term 2
(Jan. - Mar.)</th><th>Term 3
(Apr. - Jun.)</th></tr>
<tr><td>Grade 6</td><td colspan="2">The Suite Life of Zack and Cody</td><td>Wizards of Waverly Place</td></tr>
<tr><td>Grade 10</td><td rowspan="2">I Love Lucy</td><td colspan="2">Seinfeld</td></tr>
<tr><td>University</td><td colspan="2"></td></tr>
</table>

As students' motivation was considered a key factor, at the end of each term students in the EG were enquired about their opinion of the TV series they had been watching and were asked whether they preferred changing it or sticking to the same for the next term. At the end of the first term, Grade 6 students expressed their preference for watching more episodes of the same TV series, and they were willing to know what happened next. However, at the end of the second term, they showed some signs of boredom and they opted for changing the TV series. So, in consonance with their preferences, another TV series, similar in content and difficulty was used. On the other hand, Grade 10 participants chose to change the TV series after the first term and continued with a different TV series for Terms 2 and 3.

6.5.1.1. TV series at Grade 6: 'The Suite Life of Zack and Cody' and 'Wizards of Waverly Place'.

The Suite Life of Zack and Cody (Kallis, Dreayer, Eells, & Geoghan, 2005) was found to be appropriate for Grade 6 students and they watched 16 episodes along six months (Terms 1 and 2). This American sitcom portrays the adventures of a single mother (Carey) of two twins (Zack and Cody) living in the Tipton Hotel in Boston, where she sings and acts in one of its lounges. Her two sons are involved in mischiefs and they often get in trouble, sometimes unintentionally, and try to find some excuses and funny ideas to get out of it. The series also centres upon the hotel owner's daughter (London), the candy-counter girl (Maddie) and the manager (Mr Moseby), all of whom are directly involved in the twins' adventures.

The series has a total of three seasons, and it was aired on Disney Channel US between March 2005 and September 2008. These three seasons are divided into 87 weekly episodes (Season 1 - 26 ep., Season 2 - 39 ep., and Season 3 - 22 ep.). However, all episodes chosen for the study were from Season 1, originally aired in the United States between March 2005 and January 2006. As it was quite successful, it was also broadcasted internationally (Australia, New Zealand, Canada, etc.), including Spain, where it was translated as *Hotel, dulce hotel: Las aventuras de Zack y Cody*, and was first aired in September 2005 on Antena 3, a private nation-wide channel. When it first aired in Spain, children participating in the intervention were only 1 year old. Besides, Season 1 was not released in DVD in Spain, it was only released in

Germany and Greece in 2008 and 2010, respectively. The German DVD set included subtitles in many European languages (Spanish among them) and it could be used for the intervention.

Wizards of Waverly Place (Greenwald, Murrieta, Cheung, & Montanio, 2007) was also found to be suitable for this group. The series pictures the adventures of the Russo family, which is composed of five members. The three siblings (Alex, Justin and Max) and the father (Jerry) have magical powers. However, they enter a competition as only one of them will keep their magical powers forever while the others will lose them. The three young teenagers need to conceal their magical powers to the rest of the characters in the series, mainly to their friends at high school, or to the customers in the sandwich shop the family owns.

This American fantasy sitcom, targeting also old children and young teenagers, originally aired on Disney Channel US from October 2007 to January 2012. The series has a total of 106 episodes, divided into four seasons (Season 1 - 21 ep., Season 2 - 30 ep., Season 3 - 28 ep., and Season 4 - 27 ep.), although all the eight episodes selected for the intervention belonged to Season 1 (aired in the United States between October 2007 and August 2008). In Spain, the series was broadcasted with the title *Los Magos de Waverly Place* on Disney Channel, a private nation-wide channel available in Digital Terrestrial Television, between April 2008 and March 2012 (Season 1 from April 2008 to May 2009). Thus, when it was aired, Grade 6 students were between 3 and 4 years old; therefore, they could not probably have watched the show because it targeted an older audience. The complete first season was released in DVD in Germany in September 2009, with subtitles in many European languages including Spanish, so it could be used for the purpose of the dissertation.

Although the plot was different and the vocabulary included some specific words related to magic, it resembled *The Suite Life of Zack and Cody* in many aspects: both aired on the same channel, were produced by the same company (It's a Laugh Productions, Inc.) around the same time (March 2005 vs. October 2007), and targeted the same population (old children and young teenagers). They were also similar in terms of vocabulary demands (as section 6.5.1.3.1. will show).

6.5.1.2. TV series at Grade 10 and university: 'I Love Lucy' and 'Seinfeld'.

Participants at Grade 10 and university did not show large differences in terms of age, proficiency and interests, so the same TV series was used for the two groups in Term 1: both watched *I Love Lucy* (Oppenheimer & Arnaz, 1951). This sitcom was very popular in the 1950s and it is usually regarded as one of the most influential shows ever, achieving very high rates of popularity during the six seasons it lasted. The story centres upon the four main characters: Lucy Ricardo and her husband Ricky, and their two best friends and landlords Fred and Ethel Mertz. Throughout the series, Lucy tries to get into the show-business although she is not very talented. Her husband, Ricky Ricardo, plays the role of a Cuban-American singer and musician, who has to cope with Lucy's attempts to enter the music industry. The two characters often appear together with their neighbours: Ethel Mertz is a former model who tries to make her way back to show-business and, contrary to Lucy, she sings and dances quite well. She is always accompanied by her husband, Fred Mertz, who served as a soldier in World War I and is tight-fisted. He is also good at singing and dancing, and often performs together with Ethel.

This TV series had a total of 180 episodes divided in six seasons (Season 1 - 35 ep., Season 2 - 31 ep., Season 3 - 31 ep., Season 4 - 30 ep., Season 5 - 26 ep., and Season 6 - 27 ep.) originally aired between October 1951 and May 1957 in black and white. All episodes were broadcasted on CBS, a very well-known private channel in the US. The eight episodes selected for the intervention aired in Season 5, between October 1955 and May 1956. In Spain, translated as *Te quiero Lucy*, it first aired in 1958 on TVE, the public broadcasting corporation. Therefore, participants in our study could not have seen it on TV. The complete fifth season was first released on DVD in 2005, although a remastered version was released in November 2012. This latest version was bought and used in the intervention.

In the next two terms, and following Grade 10 students' demands, *Seinfeld* (David et al., 1989) was the TV series selected to be watched. It is a very well-known and highly acclaimed American sitcom from the 1990s, depicting the daily life of its main character, Jerry Seinfeld, and some colleagues of his: former girlfriend Elaine, best friend George and neighbour Kramer. Throughout nine different seasons, the show portrays their adventures, many of them quotidian and non-sense, making it a very hilarious TV series about mundane things. Jerry plays the role of a stand-up down-to-earth comedian living in New York and his apartment is usually the setting of many of the adventures involving all the characters. Elaine Benes is Jerry's former girlfriend and now friend, a bit arrogant and impulsive, although extremely humorous at the same time. Kramer is Jerry's across-the-hall neighbour and he is a character with many ups and downs, being sometimes naïve and ignorant and others smart and intellectual. He usually gets into trouble and tries to find his way out of it by playing tricks that work at the beginning but ultimately fail. Finally, George is Jerry's friend and they know each other since adolescence. He feels inferior to the rest of the characters and is usually jealous of others' achievements. He is also a passionate liar, and this causes him trouble every now and then.

Nine seasons of *Seinfeld* were broadcasted and a total of 180 episodes were shot (each season had between 5 and 24 episodes). The entire series was aired between July 1989 and May 1998 in the United States on NBC, a private nation-wide broadcasting network. The episodes selected for the intervention were part of Season 4 (Term 2) and Season 5 (Term 3), which were aired between August 1992 and May 1993, and September 1993 and May 1994, respectively. This American sitcom was broadcasted in many countries throughout the world, including Spain, where it has been aired at least three times: the first broadcasting took place in January 1998 on Canal +, a commercial television channel that was quite popular in the 1990s and 2000s; again on Canal + in August 2007 and a third and final time in July 2009 on TNT, a commercial television channel that is only available in some pay-per-view platforms. Thus, it seems highly improbable that participants were familiar with the show before the intervention, since they were nine years old when it was last aired in Spain. The nine seasons were released on Region 2 DVDs between 2004 and 2007.

Although both TV series (*I Love Lucy* and *Seinfeld*) were produced at different times (1950s and 1990s), the two were set in the United States and explain the everyday adventures of the main characters and their friends, so they depict daily life situations of American households. Both TV series can be classified as situation comedies, with all the typical characteristics of the genre. Finally, the two were highly-acclaimed, as shown by the number of episodes shot and the high rates of popularity and were also similar in terms of vocabulary demands (as section 6.5.1.3.2. will show).

6.5.1.3. Episode selection.

Specific criteria were adopted to select the most appropriate episodes for the intervention: first of all, the researcher carefully and attentively watched all episodes of the same season and noted down the vocabulary that could be unknown to participants (therefore, susceptible to be learned and tested during the intervention). Moreover, he also jotted down the number of occurrences of each of these words so as to have some idea of their frequency in the episode. At the same time, content was checked, and the researcher made sure that it was appropriate to be shown in a classroom context and it was catchy enough for students to be entertained while participating in the study. Some episodes were discarded because of a lack of challenging vocabulary (or because they contained a high number of words which were thought to be unknown, but which appeared only once and were not useful to analyse the effect of video viewing on their learning). Finally, consecutive related episodes were given priority, since research proved that serial episodes from the same season are easier to understand than unrelated episodes from the same or a different show (Rodgers & Webb, 2011).

The lexical challenge of the episodes was also analysed. In order to conduct the analysis of the lexical coverage, VocabProfile v.2 (Cobb, ongoing) was used and the BNC (British National Corpus) / COCA (Corpus of Contemporary American English) 1-25k frequency lists worked as baseline (Davies, 2009; Nation, 2012). These frequency lists combine instances of British and American English as well as written (novels, letters, journals, etc.) and spoken texts (movies, television, face-to-face conversations, telephone conversations, etc.), so they were chosen because they contain a representative selection of different types of input (not just written as in other corpora). As the corpus (i.e., BNC / COCA) from which the lists were extracted was also built with spoken texts from movies and television programmes, it was considered appropriate to analyse the transcripts of the episodes selected for the experiment.

Before entering the pruned transcripts into VocabProfile, different steps were taken:

1) All transcripts were downloaded from the Springfield! Springfield! website (https://www.springfieldspringfield.co.uk), a free online database which gives access to the written transcripts of many movies and TV episodes. The transcripts that were downloaded were doublechecked to make sure that they indeed corresponded to the actual episodes, and no mismatches were detected.
2) Transcripts were pruned so as the scripts could be easily processed, and words classified by VocabProfile. Any inconsistencies between the downloaded transcript and the spoken text heard by participants (or the written text seen in the subtitles) were edited and priority was given to the spoken text / subtitles, because it was what participants were actually exposed to during the viewing sessions.
3) Contractions were analysed as two separate words (‘don’t’ → ‘do not’) and all numbers were written down in letters to facilitate classification in VocabProfile.
4) Proper nouns were classified as 1k words for two reasons: firstly, if classified as off-list words, the lexical coverage would have been extensively affected, because there were many proper nouns (mainly vocatives) and the text would have unrealistically seemed much more difficult than what it actually was. Secondly, if excluded from the analysis, the percentage of (un)known words in the text would have been misleading, as it is usually taken for granted that viewers are familiar with the names of, for example, famous places or recurring characters.

5) Compound words, namely nouns, were analysed as single lexical items and automatically classified as off-list words by the software. It was not deemed necessary to re-classify them as these compound words were indeed more difficult and less likely to be known than the corresponding separate parts. For example, participants in Grade 6 may not be familiar with the word ‘ballroom’, although they might be familiar with ‘ball’ (i.e., the spherical object to play) and ‘room’.
6) Interjections and marginal words (e.g., ‘mmm’, ‘uhm’, ‘ah’, ‘oh’) were deleted from the transcripts. They do not typically convey much meaning and they were thought to be known by all participants, as they are quite transparent and similar in the participants’ L1s.

These steps were consistently addressed in the analysis of the TV series’ transcripts to analyse lexical coverage. After that, a corpus with the episodes viewed each term was created and input to VocabProfile in order to assess the lexical comprehension demands. In this respect, research has shown that 95% lexical coverage allows for reasonable comprehension of a written text (Laufer, 1989), while 98% seems to be needed to attain ideal comprehension (Hu & Nation, 2000). Based on this, and in line with Webb and Rodgers (2009b), these two coverage levels were inspected in the episodes’ analysis.

6.5.1.3.1. Grade 6.

In Terms 1 and 2, the first sixteen episodes of the first season of *The Suite Life of Zack and Cody* were selected, since vocabulary and content were deemed to be suitable (the first eight were shown in Term 1 and the subsequent eight in Term 2). For the reasons explained in section 6.5.1., the TV series changed in Term 3. As it was new to all students, efforts were made to include the initial episodes of the series. However, this time, the episodes were not correlative as three had to be excluded due to their content and lack of testable vocabulary. Hence, the second, seventh and ninth episodes were skipped. Table 6.4 lists the twenty-four episodes, as well as their running times:

Table 6.4
List of episodes selected in Grade 6.

Episode	Title	Running time*
G6E01	"Hotel Hangout"	21m 27s
G6E02	"The Fairest of Them All"	21m 28s
G6E03	"Maddie Checks In"	21m 28s
G6E04	"Hotel Inspector"	21m 27s
G6E05	"Grounded on the 23rd Floor"	21m 25s
G6E06	"The Prince & The Plunger"	21m 27s
G6E07	"Footloser"	21m 27s
G6E08	"A Prom Story"	21m 25s
Mean T1		**21m 27s**
G6E09	"Band in Boston"	21m 27s
G6E10	"Cody Goes to Camp"	21m 26s
G6E11	"To Catch a Thief"	21m 28s
G6E12	"It's a Mad, Mad, Mad Hotel"	21m 27s
G6E13	"Poor Little Rich Girl"	21m 27s
G6E14	"Cookin' with Romeo and Juliet"	21m 27s
G6E15	"Rumors"	21m 27s
G6E16	"Big Hair & Baseball"	21m 26s
Mean T2		**21m 27s**
G6E17	"Crazy 10-Minute Sale"	21m 44s
G6E18	"I Almost Drowned in a Chocolate Fountain"	21m 41s
G6E19	"New Employee"	21m 44s
G6E20	"Disenchanted Evening"	21m 43s
G6E21	"You Can't Always Get What You Carpet"	21m 27s
G6E22	"Curb Your Dragon"	21m 07s
G6E23	"Pop Me and We Both Go Down"	21m 43s
G6E24	"Potion Commotion"	21m 41s
Mean T3		**21m 36s**
Total		**8h 35m 59s**

Note. * Excludes opening and ending themes and credits. | G = grade / E = episode.

Regarding coverage, Table 6.5 shows the lexical demands of the episodes selected in Grade 6 in each term (see Appendix B to check the vocabulary demands of each of the twenty-four episodes). As can be seen in the table, the 95% coverage level was reached between the 2k and 3k frequency bands in Terms 1 and 2, and between 1k and 2k in Term 3. This coverage threshold has been said to be enough for successful comprehension of television programmes, in part thanks to the extra support provided by the images (Webb & Rodgers, 2009b). The ideal comprehension level of 98% was reached between the 5k and 6k levels, indicating that participants needed to have mastered the first six thousand words of the English language to reach this threshold. In terms of individual episodes, more variability was observed: there were some in which 95% coverage was achieved at the 2k level whereas in others mastery of the 4k level was needed. Regarding 98% coverage, in some episodes, knowledge of 4k words was enough to reach the threshold, but there were some episodes in which such comprehension threshold was not reached until the 11k level. These coverage levels support the fact that L1 subtitles were used in Grade 6, as the mean VS of the EG at this level was 1,750 words (if L2 subtitles had been used, TV series comprehension would have been considerably more challenging).

Table 6.5
Tokens, types, WFs and cumulative coverage of the episodes viewed each term by Grade 6 students.

	TERM 1				TERM 2				TERM 3			
Frequency band	**Tokens**	**Types**	**WFs**	**Cumulative coverage** (in %)	**Tokens**	**Types**	**WFs**	**Cumulative coverage** (in %)	**Tokens**	**Types**	**WFs**	**Cumulative coverage** (in %)
1k	18,319	1,382	911	91.11	19,920	1,527	972	91.18	21,985	1,416	912	91.71
2k	752	407	352	94.85	796	453	381	94.82	923	429	338	95.58
3k	209	138	127	95.89	221	141	125	95.83	167	110	97	96.28
4k	189	131	115	96.83	221	134	116	96.84	177	103	92	97.02
5k	163	94	83	97.64	184	108	97	97.68	130	76	63	97.56
6k	87	54	51	98.07	99	50	48	98.13	118	57	54	98.05
7k	40	25	25	98.27	51	31	30	98.36	54	29	28	98.28
8k	34	25	22	98.44	29	22	22	98.49	28	18	17	98.40
9k	41	17	17	98.64	29	23	23	98.62	41	14	14	98.57
10k	20	16	16	98.74	24	18	17	98.73	42	20	18	98.75
11-25k	108	89	87	99.23	110	85	81	99.24	107	61	58	99.19
Off-list	144	105	???	99.95	164	129	???	99.99	191	102	???	99.99
Total	**20,106**	**2,484**	**≈ 1,806**	**≈ 100**	**21,848**	**2,722**	**≈ 1,912**	**≈ 100**	**23,873**	**2,436**	**≈ 1,691**	**≈ 100**

6.5.1.3.2. Grade 10 and university.

Similar to what happened in Grade 6, in the series for Grade 10 and university students, there were more episodes available than the number needed for the intervention, so the same episode selection criteria were applied (i.e., namely, content- and level-appropriateness and presence of unknown vocabulary). However, as these TV series targeted an adult audience, there were more problems in terms of content appropriateness, with rather explicit inadequate topics and scenes, especially in *Seinfeld*. After excluding those that were not appropriate for a school context and checking that the remaining ones contained sufficient testable vocabulary, the resulting list did not contain many episodes to choose from. In Term 1, the first eight episodes of the fifth season of *I Love Lucy* could not be selected, as the plot of the eighth one was rather dull. In Term 2, seven episodes were selected (one less than in the previous term as there were just seven sessions). Similarly, although efforts were made to include the first seven consecutive episodes of *Seinfeld*'s season 4, several had to be skipped due to explicit sexual references, lack of vocabulary possibly unknown by students and their length (two special episodes lasted one hour). As students wanted to continue with *Seinfeld* for another term, and there were not many episodes left in season 4, it was decided that season 5, the natural continuation of the previous season, was the most pertinent. The same episode selection criteria were applied and, once more, it was not possible to select seven consecutive episodes, as five were skipped. Table 6.6 contains the list of the 22 episodes and their running times.

Table 6.6
List of episodes selected in Grade 10 and university.

Episode	Title	Running time*
G10E01	"Lucy Visits Grauman's"	24m 40s
G10E02	"Lucy and John Wayne"	24m 55s
G10E03	"Lucy and the Dummy"	22m 15s
G10E04	"Ricky Sells the Car"	24m 31s
G10E05	"The Great Train Robbery"	24m 40s
G10E06	"Homecoming"	25m 11s
G10E07	"Face to Face"	24m 57s
G10E08	"Nursery School"	24m 50s
Mean T1		**24m 30s**
G10E09	"The Trip (Part 1)"	21m 44s
G10E10	"The Trip (Part 2)"	21m 37s
G10E11	"The Bubble Boy"	21m 34s
G10E12	"The Cheever Letters"	22m 28s
G10E13	"The Airport"	21m 45s
G10E14	"The Movie"	21m 01s
G10E15	"The Shoes"	21m 22s
Mean T2		**21m 39s**
G10E16	"The Glasses"	21m 46s
G10E17	"The Sniffing Accountant"	21m 47s
G10E18	"The Bris"	21m 19s
G10E19	"The Lip Reader"	21m 35s
G10E20	"The Non-Fat Yogurt"	21m 42s
G10E21	"The Conversion"	21m 48s
G10E22	"The Stall"	21m 47s
Mean T3		**21m 41s**
Total university		**3h 15m 59s**
Total Grade 10		**8h 19m 14s**

Note. * Excludes opening and ending themes and credits. | G = grade / E = episode.

As seen in Table 6.7, in terms of cumulative coverage, the two TV series posed similar lexical demands. 95% lexical coverage was reached between 1k and 2k in Terms 1 and 2, whereas, in Term 3, the threshold was reached between the 2k and the 3k frequency bands. Regarding 98% coverage, in two of the terms (T1 and T2), knowledge of the first five thousand words of the English language sufficed to reach the threshold. In Term 3, when students watched *Seinfeld*'s season 5, 98% comprehension was reached at the 6k level. Focusing on each of the twenty-two episodes (see Appendix C to check the lexical demands of each of them), we see that 95% coverage could be attained with knowledge of the first two or three thousand words in all the episodes; although most of the times it was reached at the 2k level. There was a bit more variability with respect to 98% coverage as, in some of the episodes, mastery of the 3k level sufficed whereas, in others, it could not be reached without knowledge of the first eight thousand words. However, in 17 out of the 22 episodes (77.27%), the coverage threshold was reached at the 5k, 6k or 7k frequency bands. Taking into account that the VS of the EGs was 3,498 (in Grade 10) and 4,431 words (at university), the lexical coverage of the episodes selected further justifies that subtitles were presented in L2 at both levels.

In the next sections, the materials especially designed for the intervention on the basis of the TV series will be presented. First, the target vocabulary selected from each of the episodes will be described, followed by the vocabulary tests and tasks, as well as the comprehension tests.

Table 6.7
Tokens, types, WFs and cumulative coverage of the episodes viewed each term by Grade 10 and university students.

	TERM 1				TERM 2				TERM 3			
Frequency band	**Tokens**	**Types**	**WFs**	**Cumulative coverage** (in %)	**Tokens**	**Types**	**WFs**	**Cumulative coverage** (in %)	**Tokens**	**Types**	**WFs**	**Cumulative coverage** (in %)
1k	25,145	1,401	938	94.24	20,415	1,452	991	92.54	19,230	1,412	953	91.60
2k	696	359	301	96.85	702	411	344	95.72	704	420	339	94.95
3k	137	78	69	97.36	184	128	117	96.55	218	166	147	95.99
4k	149	73	67	97.92	221	93	81	97.55	195	119	105	96.92
5k	112	57	50	98.34	127	71	62	98.13	127	72	63	97.52
6k	80	44	40	98.64	83	48	42	98.51	111	61	49	98.05
7k	58	28	23	98.86	47	32	30	98.72	73	42	38	98.40
8k	26	16	15	98.96	29	19	18	98.85	56	35	33	98.67
9k	24	17	16	99.05	37	18	15	99.02	32	20	18	98.82
10k	27	17	15	99.15	30	10	10	99.16	27	22	21	98.95
11-25k	92	56	54	99.49	60	37	36	99.43	98	62	58	99.40
Off-list	137	70	???	100	126	73	???	100	123	68	???	99.99
Total	**26,683**	**2,217**	**≈ 1,588**	**≈ 100**	**22,061**	**2,393**	**≈ 1,746**	**≈ 100**	**20,994**	**2,500**	**≈ 1,824**	**≈ 100**

6.5.2. Vocabulary tests.

6.5.2.1. Target vocabulary.

6.5.2.1.1. Target vocabulary selection.

Five TWs were selected from each of the episodes, a number which was constrained by a couple of factors:

a) First, the results of the pilot study (to be explained in the 'Procedure' section) showed that more than five words were too many to include in the same viewing session. The initial idea to include eight target items was reduced to five, as this would have also implied that learners would have been tested on a total of 64 words each term, with longer tests.
b) As the TWs' selection was limited by the television series and episodes, choosing more than five TWs per episode would have implied selecting items appearing only once.

Hence, we have 35 or 40 TWs depending on whether seven or eight episodes were watched per term. The number of TWs at each of the levels for every term and the academic year can be found in Table 6.8:

Table 6.8
***N* of TWs on which participants were tested.**

Level	Term 1 (*n*)	Term 2 (*n*)	Term 3 (*n*)	Total (*N*)
Grade 6	40	40	40	**120**
Grade 10	40	35	35	**110**
University	40			**40**

Target vocabulary was selected on the basis of different criteria so that different types of words were included in the sample:

- Unknown vocabulary → Only the TWs which were thought to be unknown to students were pre-selected. The researcher attentively watched all the episodes and noted down those words which, according to his experience, would be unknown to Grade 6, Grade 10 and university participants. During this process, special attention was taken to only pre-select words that were relevant for the storyline of the episode. When this was not possible, less relevant words were also jotted down.
- Frequency in the episode → Episode transcripts were entered into VocabProfile (Cobb, ongoing) to check the actual frequency in the episode of those words pre-selected by the researcher. TWs occurring only once in the episode were discarded, since research has proved that different encounters with a lexical item are needed in order to learn it (Pellicer-Sánchez & Schmitt, 2010; Pigada & Schmitt, 2006; Rott, 1999; Uchihara et al., 2019; Waring & Takaki, 2003; Webb, 2007). Whenever possible, those TWs with a higher number of occurrences were selected; thus, words occurring five times were prioritised over words occurring four times and these over those encountered three times, etc.
- Word category and word concreteness → Although most words fulfilling the criteria were nouns and verbs, adjectives, adverbs and prepositions were also included, in order for all word categories to be represented in the sample. Furthermore, concrete words were more numerous in the sample, even though abstract words were also chosen.
- Cognateness → Cognates were understood as "words that are formally (phonologically or orthographically), semantically, and etymologically related in two languages" (Peters & Webb, 2018, p. 5). Since cognates tend to be easier to acquire than non-

cognates (Otwinowska, 2016) because they share some aspects of form and meaning with an item in the L1, they were not often selected as TWs. Presence of too many cognates would have led to an overestimation of the gains, which should have been adjusted downwards at a later stage. However, some cognates were included (as they are also found in language).

6.5.2.1.2. Target vocabulary description.

In the following paragraphs, TWs will be described according to the criteria proposed in the previous section: (1) audiovisual input frequency (measured in the episode in which the words were target and also in the corpus of seven / eight episodes that were watched in the given term); (2) frequency in the language (as measured according to the COCA Corpus and the SUBTLEX$_{US}$ Corpus); (3) PoS; (4) concreteness rating; and (5) cognateness (see Tables 6.9-6.22 for the features of the TWs). Even though some of these descriptors are transparent, others require some explanation. Further information is also given to justify why we selected the corpora / lists on which our description is based:

- The COCA Corpus (Davies, 2009) contains more than 560 million words from 1990 to the present year. It covers a wide range of sources including spoken and written texts (the latter taken from fiction books, magazines, newspapers and academic journals). The corpus was considered adequate to describe our TWs because it thoroughly reflected discourse in media and audiovisual materials as it contains more than 79 million words taken from "transcripts of unscripted conversation from more than 150 different TV and radio programs" (Davies, 2009, p. 161). As mentioned by the author, it is updated every year and, in December 2017, it already counted with more than

220,000 texts, equally balanced between the different genres and modes. The author considers that even if the only spoken texts included in the corpus come from unscripted television, this does not threaten its validity, as they are a reflection of spontaneous, at-home, non-media conversation. Finally, there is also a lack of Internet texts, although the author claims that it was not possible to compile the same number of texts (and thus words) from 1990 to the present year, namely because Internet was not such a popular resource back in the early 1990s.

- Another corpus that was used to describe the TWs was the $\text{SUBTLEX}_{\text{US}}$ Corpus (Brysbaert & New, 2009). It now contains more than 8,300 files from different sources: US films from 1900-1990, 1990-2007 and television series. In total, the corpus counts with 51 million words (16.1 from TV series, and 34.9 from American films). It was considered a very good tool for the doctoral dissertation, since it is the only corpus we are aware of that is exclusively built on subtitled audiovisual materials. Besides, the authors claim that television and movies have a solid basis for being recognized as natural spoken texts as they reflect people's behaviour in social interactions (New, Brysbaert, Veronis, & Pallier, 2007). When compiling the corpus, the authors restricted their search to subtitles from American films and television programmes, and thus discarded any materials from other English-speaking areas. This made the corpus even more relevant to the present study because all the TV series used in the intervention were American. Different measures were calculated on the basis of the corpus, but only one will be reported in the dissertation: SUBTL_{CD}, which indicates the percentage of films and television series in which the TWs appear. It can go from 0 (very uncommon) to 100 (very frequent), indicating that, for instance, the word 'look' (with a value of 98.83) is much more common in American television subtitles than the word 'jargon', with a value of 0.33. The measure has two decimal values to ensure precision and to convey as much information as possible.

- Information on concreteness is provided using a 5-point scale (Brysbaert, Warriner, & Kuperman, 2014), in which 1 means totally abstract (language-based) and 5 fully concrete (experience-based). The authors conducted a study in which 30 residents in the US rated an initial sample of 60,099 words and 2,940 two-word combinations taken from the SUBTLEX$_{US}$ Corpus. However, as raters were also asked about their familiarity with the words, only those words known by at least 85% of them were included in the final analysis. This left the authors with a final list of 39,954 items (37,058 words and 2,896 two-word expressions). In the instructions given to participants, concrete words were defined as follows: "a concrete word comes with a higher rating and refers to something that exists in reality; you can have immediate experience of it through your senses (smelling, tasting, touching, hearing, seeing) and the actions you do" whereas an abstract item was defined as a word which "comes with a lower rating and refers to something you cannot experience directly through your senses or actions. Its meaning depends on language" (Brysbaert et al., 2014, p. 906). For description purposes, only the mean concreteness rating will be reported in the tables.

Table 6.9
***N* of encounters with the TWs in the episodes in which they were target in Grade 6.**

Word frequency (*N* repetitions episodes)	*N*
1	-
2	21
3	24
4	23
5	20
6	8
7	3
8	1
9	6
10	-
11	2
12	3
13	2
14	2
15	-
16	2
17	1
18-22	-
23	1
24-27	-
28	1
Total	**120**

Table 6.10
***N* of encounters with the TWs in the term in which they were target in Grade 6.**

Word frequency (*N* repetitions term)	*N*
1	-
2	16
3	20
4	18
5	19
6	8
7	2
8	2
9	9
10	3
11	6
12	4
13	-
14	2
15	1
16	2
17	1
18	2
19	1
20-22	-
23	1
24-27	-
28	1
29	1
30-46	-
47	1
Total	**120**

Table 6.11
Frequency of the TWs tested in Grade 6 according to the COCA corpus.

Word frequency (according to COCA)	*N*
1k	21
2k	27
3k	10
4k	12
5k	14
6k	6
7k	3
8k	-
9k	4
10k	5
11k	-
12k	2
13k	2
14k	-
15k	1
16k	-
17k	1
18k	1
19k	-
20k	1
21k-25k	-
Off-list	10
Total	**120**

Table 6.12
Frequency of the TWs tested in Grade 6 according to the $\text{SUBTLEX}_{\text{US}}$ corpus.

Word frequency (according to $\text{SUBTLEX}_{\text{US}}$; in %)	*N*
0-5	66
5.01-10	27
10.01-15	6
15.01-20	4
20.01-25	4
25.01-30	3
30.01-35	2
35.01-40	3
40.01-45	1
45.01-50	1
50.01-60	-
60.01-65	1
65.01-80	-
80.01-85	1
85.01-90	-
90.01-95	1
95.01-100	-
Total	**120**

Table 6.13
Part of speech of the TWs tested in Grade 6.

Part of speech	*N*
Adjectives	13
Adverbs	1
Nouns	77
Prepositions	1
Verbs	28
Total	**120**

Table 6.14
Degree of concreteness of the TWs tested in Grade 6.

Concreteness (mean)	*N*
1-2	7
2.01-3	23
3.01-4	27
4.01-5	62
Not in the list	1
Total	**120**

Table 6.15
Cognate status of the TWs tested in Grade 6.

Cognateness (Y/N)	*N*
Cognates	14
Non-cognates	106
Total	**120**

Table 6.16
***N* of encounters with the TWs in the episodes in which they were target in Grade 10 / university.**

Word frequency (*N* repetitions episodes)	***N***
1	-
2	27
3	20
4	22
5	8
6	10
7	4
8	7
9	3
10	-
11	3
12	2
13	2
14	1
15	-
16	1
Total	**110**

Table 6.17
***N* of encounters with the TWs in the term in which they were target in Grade 10 / university.**

Word frequency (*N* repetitions term)	***N***
1	-
2	18
3	24
4	21
5	9
6	11
7	7
8	5
9	4
10	-
11	3
12	2
13	3
14	1
15-19	-
20	1
21-26	-
27	1
Total	**110**

Table 6.18
Frequency of the TWs tested in Grade 10 / university according to the COCA corpus.

Word frequency (according to COCA)	***N***
1k	6
2k	16
3k	6
4k	16
5k	10
6k	12
7k	6
8k	5
9k	7
10k	5
11k	2
12k	3
13k	3
14k	1
15k	1
16k	3
17k	2
18k-25k	-
Off-list	6
Total	**110**

Table 6.19
Frequency of the TWs tested in Grade 10 / university according to the SUBTLEX$_{US}$ corpus.

Word frequency (according to SUBTLEX$_{US}$; in %)	***N***
0-5	83
5.01-10	15
10.01-15	4
15.01-20	-
20.01-25	2
25.01-30	1
30.01-35	-
35.01-40	1
40.01-50	-
50.01-55	1
55.01-60	1
60.01-100	-
Not in the list	2
Total	**110**

Table 6.20
Part of speech of the TWs tested in Grade 10 / university.

Part of speech	***N***
Adjectives	9
Adverbs	-
Nouns	71
Prepositions	-
Verbs	30
Total	**110**

Table 6.21
Degree of concreteness of the TWs tested in Grade 10 / university.

Concreteness (mean)	***N***
1-2	3
2.01-3	8
3.01-4	31
4.01-5	59
Not in the list	8
Total	**110**

Table 6.22
Cognate status of the TWs tested in Grade 10 / university.

Cognateness (Y/N)	***N***
Cognates	12
Non-cognates	98
Total	**110**

As can be seen in Tables 6.9-6.15, the TWs selected in Grade 6 had different features and covered a wide range of the vocabulary that learners can find in many communicative situations. In terms of repetitions in the episodes in which TWs were the object of study, 26.67% were repeated six or more times. From Table 6.10, it can be seen that most of the TWs appeared in some of the other episodes of the term when these were tested, although the number of encounters was not much higher. Regarding their frequency of occurrence in the language, data from the COCA Corpus shows that most of the TWs belonged to the first, second, third, fourth and fifth one-thousand-word frequency bands (70%), further indicating that they were quite frequent in the English language, what Nation (2013) labelled as high- and mid-frequency words. However, it is also true that most of them were not typical of American film and television subtitles: 77.5% were classified in the first ten percent of the $\text{SUBTLEX}_{\text{US}}$ corpus, which is the same as saying that they are one of the 7,429 most infrequent words used by the audiovisual American industry. Regarding PoS, 64.17% of the TWs were nouns, followed by verbs (23.33%), adjectives (10.83%) and adverbs and prepositions (1.67%). In terms of concreteness, most of the target items were classified as being rather concrete by the raters in Brysbaert et al. (2014), with a mean rating of 3.83 out of five. Finally, a high percentage of the TWs (88.33%) were not cognates between English and Catalan / Spanish. For more information on the TWs selected in Grade 6, please check Appendix D, where all the TWs and their specific traits are listed.

Regarding the vocabulary selected in Grade 10 and at university, as seen in Tables 6.16-6.22, 30% of the TWs appeared six or more times in the episode they were target, and the same applies to their frequency in the term in which they were tested. These tables also show that some of the TWs were encountered in other episodes, although the number of repetitions was not very high. Most of the TWs belonged to the second, fourth, fifth and sixth frequency bands in the COCA Corpus (49.09%), that is, most of them were high- or mid-frequency words. Nevertheless, 89.09% of the TWs were classified in the first ten percent of the $\text{SUBTLEX}_{\text{US}}$ corpus, showing that they were not very common in American films or television series. Regarding the TWs' grammatical category, 64.55% were nouns, 27.27% were verbs and 8.18% were adjectives; there were no adverbs and prepositions. Most of the lexical items (53.64%) were classified as being rather concrete by Brysbaert et al. (2014) and the average of the TWs' concreteness rating was 4.01 points out of five. Finally, only 10.91% of the selected words were considered to be English cognates in Spanish or Catalan, whereas 89.09% were non-cognates. For more information on the TWs selected in Grade 10 and university, please check Appendix E.

As can be noted, the characteristics of those TWs tested in Grade 6 and Grade 10 / university were similar. For instance, at both levels, around 30% of the TWs were encountered six or more times in the episodes they were target. In addition, most of them were considered to be high- or mid-frequency words according to Nation (2013) and rather infrequent in American films or television series. Besides, in both Grade 6 and Grade 10 / university, a higher number of nouns and verbs were tested (as opposed to adjectives, adverbs and prepositions), the majority of which being concrete and non-cognates.

6.5.2.2. Vocabulary pre- and post-tests.

Pre- and post-tests assessed all the TWs for a given term (35 or 40 depending on the number of sessions). As there were too many TWs to do the whole test in one go and it could have been too tiring for learners, who were not used to this test format, it was divided into two parts: in the first part, twenty TWs were included (those from episodes one to four) while the second part included 15 or 20 TWs (those from episodes five to seven / eight). The order in which the TWs appeared on the tests was quasi-randomly established by the 'Sequence Generator' function on the RANDOM.ORG website (www.random.org) (Haahr & Haahr, 1998), which randomizes a sequence of one's choice. Once the order was established by the software, it was revised by the researcher so as to avoid having two words with similar forms too close to one another (for instance, if 'to snub' and 'stub' appeared together, the order would have been altered and some other words would be put between them). Besides, four practice items were added at the beginning of the test in order to familiarize participants with the type of test and to avoid that lack of familiarity affected the scores. The four items were usually easy words (e.g., 'to drink' or 'paper') although some more difficult items (e.g., 'to plummet') were incorporated to 'warn' students about the difficulty of the test they were about to take.

The learners' task was to listen to an audio file recorded by an American NS, write down the English form of the word and provide the Spanish / Catalan translation in case they knew it. English synonyms or definitions in English, Spanish or Catalan were also accepted. This format made possible the calculation of three different scores:

- Knowledge of TW forms.
- Knowledge of TW meanings.
- Knowledge of both forms and meanings of the TWs.

Time was given to students to complete and revise their answers at the end of the tests. Sufficient space was also provided for them to write all their answers and more space was available on the two A4 pages on which the tests were printed. The vocabulary pre- and post-tests are included in Appendix F.1.

In the audio file, each TW form was repeated twice, with a five-second pause between both repetitions, followed by a ten-second interval preceding the next item. All the verbs were presented in their base to-infinitive form. Furthermore, all the TWs were numbered to prevent that any students got lost and to secure that answers always referred to the correct target item. The duration of the audio mostly depended on the number of TWs that were tested (35 or 40); their duration can be checked in Table 6.23:

Table 6.23
Duration (in minutes and seconds) of the audio files used for the pre- and post-tests.

Level	Term 1		Term 2		Term 3	
	Part I	Part II	Part I	Part II	Part I	Part II
Grade 6	5m 11s	5m 15s	5m 02s	5m 06s	5m 58s	5m 49s
Grade 10 / University	5m 12s	5m 20s	5m 06s	3m 45s	5m 37s	4m 35s

All vocabulary pre- and post-tests followed the same format. Concurring with what research suggests (Nation & Webb, 2011), different aspects of the TWs were tested. The tests were designed to assess written form recall (prompted by the aural form of the words) and meaning recall. Hence, they tapped into different types of knowledge: receptive knowledge of the spoken form, productive knowledge of the written form, and knowledge of meaning (Webb, 2007). The format that was used probably is one of the most challenging since production, and not just recognition, was demanded to learners. This format excluded the possibility to analyse partial knowledge of the TWs, although this was controlled for in the scoring phase, when three different scoring criteria were applied.

6.5.2.3. Vocabulary delayed test.

The delayed tests assessed those TWs that learners had been exposed to during Term 3. This implies that Grade 6 participants (now in Grade 7) were tested on 40 TWs and Grade 10 learners (now in Grade 11) on 35, since one session had to be cancelled in Term 3 in this case. TWs from Terms 1 and 2 were not included since participants last encountered them one year and two months before the administration of the delayed test. As chances were that learners would not remember them, it was decided to only include the TWs last encountered in June 2016, eight months before the vocabulary delayed test was taken. In addition, assessing 110 or 120 words would have taken too long for the time we had available at school.

It followed the same format as the pre- and post-tests that were used in previous terms (see section 6.5.2.2.). By using the same test format, it was possible to see which TWs were retained or not. Once more, students were asked to listen to the same audio file that had listened to eight months before, write down the English forms of the items and provide the Spanish / Catalan translation (or a synonym or a definition) in case they knew it. The only difference was that the test was not divided into two parts, and only one longer audiofile was used (11m 34s in Grade 6 and 10m 02s in Grade 10). Four practice items were also included so as to remind students how to do the test.

6.5.3. Vocabulary and comprehension tasks.

Vocabulary pre- and post-tasks aimed at introducing the TWs to students and at increasing the number of encounters students had with the TWs in each session. These vocabulary tasks were especially tailored on the basis of each episode and its format was also discussed with school teachers so that tasks were relevant for students and adequate for their proficiency level. Pre-tasks were always done at the beginning of each viewing session whereas post-tasks were administered at the end. While-watching tasks, as their name indicates, were done while viewing the TV series and they were comprehension exercises. In the next three subsections, the main traits of each type of task will be explained.

6.5.3.1. Vocabulary pre-task.

All vocabulary pre-tasks were aimed at introducing the five TWs to participants using different formats in order not to be too repetitive. They consisted in one activity through which learners became familiar with the target items and students could ask questions regarding the TWs' forms and meanings (see Appendix F.2.). All of them followed a focus-on-forms approach, since students practised several vocabulary items in a non-communicative situation (i.e., non-authentic language tasks) and paid attention to the words because they were the objects of study (Laufer, 2006). This approach was chosen, as opposed to others like focus-on-form (Long, 1998) or focus-on-meaning (Norris & Ortega, 2000), because we wanted participants to be familiarised with forms and meanings of the TWs in a very short period of time, and this was less likely to happen in a more communicative approach. In addition, research has shown that focus on forms is more effective than focus on form in terms of vocabulary learning (Laufer, 2006).

There were seven types of activities proposed, which varied across episodes:

1) Crosswords in which learners could check the TWs' definitions in case they needed help. Some letters in the crosswords were already provided.
2) Exercises in which learners had to fill in some sentences with the TWs. In some exercises, learners could check the definitions while in others this was not possible.
3) Activities in which learners had to match TWs with their definitions.

4) Exercises in which participants were provided with a visual representation of the TW and had to match this to the written form.
5) Word-searches in which the definition of the TWs was provided, as well as the number of letters the TWs had. The first letter was also circled in the word-search.
6) Exercises in which participants were instructed to order the TWs' graphemes and then fill-in sentences with them.
7) Clozes in which participants were given the first letter of the TWs and their definition and had to provide the rest of the form.

These activities were distributed along the year so that none of them was repeated too often. However, easier tasks (e.g., matching images with the TWs) were more common in Grade 6, while tasks in Grade 10 / university were slightly more challenging (e.g., cloze tests). The distribution of these exercises throughout the intervention was as follows:

Table 6.24
Distribution of vocabulary pre-tasks depending on participants' level.

Level	Cloze tests	Crosswords	Fill-in-the-gaps	Matching pictures with TWs	Matching TWs with their definitions	Ordering graphemes and fill-in-the-gaps	Word-searches	Total
Grade 6	-	3	6	10	1	1	3	**24**
Grade 10	5	3	6	2	1	-	5	**22**
University	1	1	2	-	1	-	3	**8**

6.5.3.2. While-watching task.

The while-watching task aimed at keeping participants' attention on the video they were watching, following Rodgers (2013). It consisted of two detailed multiple-choice questions, with three options each, where just one of the options was correct (see Appendix F.3.). The position of the key was randomly determined by a sequence generator and the two distractor answers were placed in the other two positions. In order to measure participant's comprehension of the first half of the episode, and not their L2 proficiency level, this task was presented in one of the learners' L1s (i.e., Spanish) (Nation, 2001).

Although the video was not very long (the average length being 21m 30s for Grade 6 episodes, 22m 37s for Grade 10 and 24m 30s at university), it was important to tell students that they had to watch the series attentively. In the middle of the episode, the teacher or the researcher stopped the video and learners were given the while-watching task (see Appendix G for the times when the video was paused).

Two sample questions (one for Grade 6 and one for Grade 10 / university) are:

<u>Episode G6E02</u>

P1 - ¿Cómo consiguió Tim su nueva peluca?

A) La compró en una tienda. / B) La alquiló. / C) Se la regalaron.

Q1 - How did Tim get his new wig?

A) He bought it at a store. / B) He rented it. / C) Somebody gave it to him.

<u>Episode G10E07</u>

P1 - ¿Qué quiere comprar Lucy al saber que su casa va a salir por televisión?

A) Una televisión nueva. / B) Un sofá nuevo. / C) Unas cortinas nuevas.

Q1 - What does Lucy want to buy when she discovers their apartment will appear on TV?

A) A new television set. / B) A new coach. / C) New curtains.

6.5.3.3. Vocabulary post-task.

In the vocabulary post-task, only the five TWs that were introduced in the pre-task were included. Participants were asked to listen to an audio of the same NS who recorded the pre- and post- vocabulary tests. Students listened to each of the five TWs' forms twice; with a five-second pause between repetitions and a ten-second interval between items. Learners were instructed to write down the English form of the words in the spaces given and then select the Spanish translation out of six possible answers. Hence, the task was a written form recall (prompted by the aural form of the words) and a meaning recognition exercise, and it analysed students' receptive knowledge of the TWs' spoken forms, their productive knowledge of the TWs' written forms, and their recognition of the TWs' meanings. The six options were provided in Spanish because "the use of the first language to convey and test word meaning is very efficient" (Nation, 2001, p. 351) and were adapted from Rodgers (2013). These were:

1) The key.
2) A semantically related distractor from the same PoS as the TW.
3) A distractor with phonological similarities with the English form of the target item. Whenever possible, it also shared the same PoS as the TW.
4) A distractor from the same frequency band and different PoS as the TW.
5) A hapax, an item which appeared in the episode although it was not selected as a TW.
6) An 'I don't know' option.

The final order of the options for each item was determined in the following way: the key was randomly placed in one of the first five positions as the last one was always reserved for the 'I don't know' option. This last option was included to minimize guessing effects, since students were explicitly told to select it if they were not completely sure about the answer (Zhang, 2013). The key's exact position was determined by the 'Sequence Generator' function in the RANDOM.ORG website (Haahr & Haahr, 1998). Distractors #2 and #3 were placed in the two closest positions to the key. In those cases when the key was in the first or last position (options A and E), distractors #2 and #3 were alternated so that a balanced number of questions had the same type of distractor next to the key. In these cases, the second closest available position was reserved for the distractor that could not be placed next to the key. Once two positions were already filled, distractor #5 was placed in the closest available position to the key. Finally, distractor #4 was placed last in the remaining available position. Examples of questions of vocabulary post-tasks administered to Grades 6 and 10 students are shown below (for all the rest, see Appendix F.4.):

Grade 6 - TW: Bossy (episode G6E09, item #2)

a) Copiar (distractor #4) [to copy]
b) Lavandería (distractor #5) [laundry]
c) Café (distractor #3) [coffee]
d) Perjudicial (distractor #2) [harmful]
e) Mandón (key) [bossy]
f) No lo sé ('I don't know' option)

Grade 10 - TW: To melt (episode G10E20, item #2)

a) Repugnante (distractor #4) [disgusting]

b) Solidificar (distractor #2) [to solidify]

c) Derretirse (key) [to melt]

d) Cinturón (distractor #3) [belt]

e) Empate (distractor #5) [tie]

f) No lo sé (‘I don’t know’ option)

6.5.4. Comprehension test.

The format of the comprehension test for each episode was adapted from Rodgers (2013) and comprised T/F, MC and ordering questions. More precisely, the tests always consisted of five T/F questions, five MC questions and eight statements that learners had to order chronologically (see Appendix F.5. for all the comprehension tests designed for the study). To avoid measuring reading skills, instead of content comprehension, the comprehension tests were administered in Spanish, one of the participants’ L1s. Another very important aspect that was considered when designing the comprehension tests was that the questions could not be answered without having watched the episode; that is, learners could not rely solely on logic to answer them. In addition, learners had to understand the content of the aural (audio) or written text (subtitles) in order to answer the questions properly; they could not answer correctly just from seeing the images.

Questions in the test could ask about detailed, general or inferential information. Detailed questions measured the participants’ ability to understand very specific information presented on the episodes. They enquired about details which were not crucial to follow the main plot.

General questions asked about the storyline of the episode or lengthy sequences, but they did not go into detail in any of the scenes. These aimed at finding out whether learners had made sense of the episode as a whole. Finally, inferential questions were included to measure participants' ability to connect different aspects of the plot or to understand some of the events that happened in the episode although they were not very explicitly shown or explained. A fairly balanced number of questions from each type was included although it was not always possible to test the same proportion of question types. Table 6.25 presents the percentage of each type of question in both the T/F and MC exercises:

Table 6.25
Percentage of detailed, general and inferential questions included in the comprehension tests.

Question type (in %)	**Grade 6**				**Grade 10 and university**			
	Term 1	**Term 2**	**Term 3**	**Mean**	**Term 1**	**Term 2**	**Term 3**	**Mean**
Detailed	45	38.75	30	**37.92**	52.50	37.14	25.71	**38.45**
General	33.75	40	35	**36.25**	36.25	41.43	51.43	**43.04**
Inferential	21.25	21.25	35	**25.83**	11.25	21.43	22.86	**18.51**

As can be seen in Table 6.25, there is always a combination of detailed and general questions. At both levels, the amount of inferential questions was lower in comparison to the other two question types due to the content and length of the episodes, which did not allow to have many of this type.

Regarding the format of the exercises, T/F and MC questions were chosen since learners were very familiar to them. They usually find these types of questions in regular class activities and exams, and they did not require much explanation. Both T/F and MC questions have the advantage that their scoring is objective, dichotomous (right or wrong) and time-saving. Furthermore, T/F questions do not require much time to be answered and they can gather large amounts of information (Rodgers, 2013): as viewing sessions lasted 50-55 minutes, and many tasks had to be done (vocabulary pre-task, episode viewing, while-watching task, vocabulary post-task and comprehension test), time was indeed a constraint. T/F questions also prompt different types of comprehension, among them literal and inference comprehension (Day & Park, 2005), two of the three constructs that the tests tapped into. As suggested by Cain and Oakhill (2006), T/F questions pose low processing demands to test takers, as they do not require a complex and elaborate response. For all these reasons, they were considered essential to be included in the tests.

Regarding the MC exercise, each question had three options, including the key and the two distractors. The number of options was limited to three since adding an extra distractor would not have increased test validity: research has shown that 3-option questions do not affect item discrimination and reliability, and they also make the test a bit easier in comparison to 4- and 5-option questions (Rodriguez, 2005). Furthermore, adding extra options to the questions may have resulted in having to add implausible distractors that would have been easily discarded by participants and that would have threatened the tests' discriminatory power and reliability. Caution was taken to make sure that distractors, keys and questions were not clues for other questions and did not uncover much of the content shown in the episode.

Finally, the last exercise of the comprehension test was a task in which learners had to order eight statements according to what had happened in the episode. The objective of such exercise was to measure participants' ability to process, follow and understand the video as a whole, taking into account not only single scenes but also the entire storyline (Rodgers, 2013).

In order to create the comprehension tests, the transcripts of all the episodes were divided into idea units[9] by the researcher. Caution was taken to make sure that only one question was asked for each idea unit, that is, if one idea unit had been addressed in a T/F question, it did not appear again in the MC or ordering exercise, and vice versa. An example of an idea unit which was transformed into a T/F question is the following: in one of the episodes of *Wizards of Waverly Place*, a mother wants her daughter to spend the summer with her Spanish grandmother because she does not speak proper Spanish. This idea unit could be turned into a T/F question (false in this case) in the following way:

Episode G6E18 - Exercise 1, Item #3

T / F Theresa cree que tendría que enviar a Alex a un campamento de lenguas porque así mejoraría su nivel de español.

T / F Theresa thinks that she should send Alex to a summer camp to improve her daughter's Spanish.

In contrast, when the idea unit was the result of more extended discussion and interaction between the characters, and several plausible answers could be developed, it was included as an MC question and two distractors were created. One of them could be easily discarded since it bared no relationship with the episode whereas the other was based on an immediately anterior or posterior event also shown in the episode, and it was considered tougher. The key's allocation was randomly determined by the 'Sequence Generator' option available in the RANDOM.ORG website (Haahr & Haahr, 1998) and the two distractors were inserted in the remaining positions. When the software determined that the key would have to be placed in options A or C, the 'tough' distractor was placed right next to the key (i.e., in option B). To

[9] Defined by Chafe (1994) as meaningful, semantically integral chunks of discourse. Ellis and Barkhuizen (2005) also claim that an idea unit is "a message segment consisting of a topic and comment that is separated from contiguous units syntactically and / or intonationally" (p. 154).

serve as an example, in an episode of *Seinfeld*, one of the characters (Newman) is worried about what his boss thinks about him, but he cannot overhear what the boss says because his boss works behind a glass. For this reason, he asks a friend of his (Laura), who is very good at lip reading, to accompany him to his workplace and read his boss' lips. This idea unit, quite lengthy, could be converted into a MC question with two distractors:

Episode G10E19 - Exercise 2, Item #4

¿Por qué Newman quiere que Laura le acompañe al trabajo un día?

A) Porque quiere hacerse el importante en el trabajo.
B) Porque quiere saber qué dice su jefe de él.
C) Porque se la quiere presentar a un compañero sordo.

Why does Newman want Laura to accompany him to his workplace?

A) Because he wants to show off in his workplace.
B) Because he wants to know what his boss thinks of him.
C) Because he wants to introduce her to a deaf workmate.

Regarding the sequencing exercise, eight different idea units were selected. The selection criterion for this last section was based on how prominent the events were for the general storyline of the episode and only events clearly taking place and verbalized were selected, omitting minor events that were less relevant and harder to spot. Having chosen the eight events, these were randomized with the same function used throughout the design of the instruments. The positions of the first and fifth events were indicated to give learners a reference point, and to avoid a null score if one statement was misplaced. Hence, participants were asked to order six of the eight statements and only these were scored. The first sentence always started with 'El episodio empieza con...' [*The episode begins with...*] followed by the first event, which took place immediately after the episode's onset. The fifth event was chosen to be indicated because it usually took place in the middle of the episode and it represented an opportunity for test takers to restart the sequencing if the order of the first half was wrong. An example of a sequencing exercise taken from *The Suite Life of Zack and Cody* is:

Episode G6E14 - Exercise 3

1	El episodio empieza con Ilsa yendo al Hotel Tipton.
________	Zack y Cody se pelean en la cocina del Hotel.
________	Descubrimos que el padre de Todd tiene un hotel.
5	London da a Todd un mechón de su pelo.
________	Todd recita poemas de Shakespeare.
________	London y Todd se besan por primera vez.
________	Maddie avisa a London y Todd de que Ilsa les está buscando en el Hotel.
________	Todd rechaza ir a Zúrich.

1	*The episode begins with Ilsa going to the Tipton Hotel.*
________	*Zack and Cody fight in the Hotel's kitchen.*
________	*We discover that Todd's father owns a hotel.*
5	*London gives Todd one of her locks of hair.*
________	*Todd recites poems by Shakespeare.*
________	*London and Todd kiss for the first time.*
________	*Maddie warns London and Todd that Ilsa is looking for them in the Hotel.*
________	*Todd rejects going to Zurich.*

6.5.4.1. Comprehension test reliability measures.

For the first two exercises of each comprehension test (T/F and MC), an item analysis was conducted to investigate test's reliability and see the "contribution it is [each item was] making to the test overall picture of candidate's ability emerging from the test" (McNamara, 2000, p. 60). The analysis could not be conducted with the last exercise, as it was an event ordering task.

In the first two exercises, difficulty and discrimination indices were calculated for each item. The difficulty index calculates the ratio of people who got the item right or wrong in comparison to the whole group and helps in deciding if test items are adequate for the test takers' level. Ideally, a test should include a combination of easy and difficult items; if many items are too easy or too difficult, all or none of the participants will get them right or wrong and it will not discriminate between more and less skilled learners. Although the ideal value for this index would be 0.50, a range from 0.30 to 0.70 is often accepted as a well-designed item (Hopkins & Antes, 1978).

On the other hand, discrimination analysis provides useful information on whether individual test items give consistent information on candidates' abilities. In doing so, the performance of high-achievers and low-achievers is compared (McNamara, 2000). In the light of this index, any value equal or higher than 0.40 is considered to be a good discriminator (del Rincón, Arnal, Latorre, & Sans, 1995). In the following paragraphs, the formulae to compute these indices and the reliability results for our tests will be presented.

6.5.4.1.1. Difficulty index.

In order to calculate the difficulty index, responses to T/F and MC questions for each test and from all test takers were entered dichotomously into statistical software (0 = wrong answer and 1 = right answer) and frequencies were calculated. For the T/F exercise, the number of students answering correctly was divided by the total number of participants who answered that question:

$$\text{Difficulty index T/F} = \frac{N \text{ of correct answers}}{N \text{ of test takers}}$$

For the second exercise (MC), another formula which controlled for the number of options that participants could choose from was needed:

$$\text{Difficulty index MC} = \frac{A - \frac{E}{n - 1}}{N}$$

where: A = *N* of correct answers; E = *N* of incorrect responses; n = *N* of options in the questions (always three); and N = *N* of test takers.

Once the difficulty index for each test item was obtained, the average for the whole test was calculated (bearing in mind the two different formulae used). In order to interpret the results, we refer to Hopkins and Antes' scale:

Table 6.26
Difficulty index values according to Hopkins and Antes (1978).

Value	Difficulty
≥ 0.86	Very easy
0.71 – 0.85	Easy
0.30 – 0.70	Moderate
0.15 – 0.29	Difficult
≤ 0.14	Very difficult

As can be seen in Table 6.27, many of the difficulty indices were around the ideal value of 0.50 (McNamara, 2000). The majority of the comprehension tests were of medium difficulty for Grade 6 learners. Out of the 24 comprehension tests, only two were found to be easy, and the majority (91.66%) were moderate. As expected, comprehension tests turned out to be a bit easier for Grade 10 students, possibly due to their higher proficiency level and more developed reading and listening skills. At this level, 59% of the tests were of moderate difficulty, and 36% were easy, and only one was found to be very easy. Finally, the three-year difference between high-school and university students, who had taken the same tests in Term 1, was also observed in the difficulty index, as 75% of the tests turned out to be easy and only 2 (25%) were of medium difficulty. Overall, these results show that the researcher-designed comprehension tests were suitable and reliable for the purpose of the study as none of them proved to be ‘difficult’ or ‘very difficult’. Besides, the fact that the tests were ‘easy’ and ‘moderate’ also shows that the TV series were adequately chosen and that learners were exposed to comprehensible input, which provides more opportunities for language learning (Krashen, 1985).

Table 6.27
Difficulty index of the comprehension tests administered to Grade 6, Grade 10 and university participants.

<table>
<tr><th colspan="4">Grade 6</th><th colspan="4">Grade 10</th><th colspan="4">University</th></tr>
<tr><th>#</th><th>Episode</th><th>Value</th><th>Difficulty</th><th>#</th><th>Episode</th><th>Value</th><th>Difficulty</th><th>#</th><th>Episode</th><th>Value</th><th>Difficulty</th></tr>
<tr><td>1</td><td>G6E01</td><td>0.60</td><td rowspan="6">Moderate</td><td>1</td><td>G10E01</td><td>0.63</td><td>Moderate</td><td>1</td><td>G10E01</td><td>0.73</td><td rowspan="4">Easy</td></tr>
<tr><td>2</td><td>G6E02</td><td>0.69</td><td>2</td><td>G10E02</td><td>0.88</td><td>Very easy</td><td>2</td><td>G10E02</td><td>0.83</td></tr>
<tr><td>3</td><td>G6E03</td><td>0.60</td><td>3</td><td>G10E03</td><td>0.79</td><td>Easy</td><td>3</td><td>G10E03</td><td>0.79</td></tr>
<tr><td>4</td><td>G6E04</td><td>0.64</td><td>4</td><td>G10E04</td><td>0.68</td><td rowspan="4">Moderate</td><td>4</td><td>G10E04</td><td>0.74</td></tr>
<tr><td>5</td><td>G6E05</td><td>0.63</td><td>5</td><td>G10E05</td><td>0.66</td><td>5</td><td>G10E05</td><td>0.70</td><td>Moderate</td></tr>
<tr><td>6</td><td>G6E06</td><td>0.60</td><td>6</td><td>G10E06</td><td>0.64</td><td>6</td><td>G10E06</td><td>0.77</td><td>Easy</td></tr>
<tr><td>7</td><td>G6E07</td><td>0.72</td><td>Easy</td><td>7</td><td>G10E07</td><td>0.52</td><td>7</td><td>G10E07</td><td>0.68</td><td>Moderate</td></tr>
<tr><td>8</td><td>G6E08</td><td>0.64</td><td>Moderate</td><td>8</td><td>G10E08</td><td>0.76</td><td>Easy</td><td>8</td><td>G10E08</td><td>0.75</td><td>Easy</td></tr>
<tr><td>9</td><td>G6E09</td><td>0.81</td><td>Easy</td><td>9</td><td>G10E09</td><td>0.59</td><td rowspan="2">Moderate</td><td colspan="4" rowspan="16"></td></tr>
<tr><td>10</td><td>G6E10</td><td>0.56</td><td rowspan="15">Moderate</td><td>10</td><td>G10E10</td><td>0.61</td></tr>
<tr><td>11</td><td>G6E11</td><td>0.56</td><td>11</td><td>G10E11</td><td>0.74</td><td>Easy</td></tr>
<tr><td>12</td><td>G6E12</td><td>0.51</td><td>12</td><td>G10E12</td><td>0.69</td><td rowspan="3">Moderate</td></tr>
<tr><td>13</td><td>G6E13</td><td>0.43</td><td>13</td><td>G10E13</td><td>0.50</td></tr>
<tr><td>14</td><td>G6E14</td><td>0.57</td><td>14</td><td>G10E14</td><td>0.55</td></tr>
<tr><td>15</td><td>G6E15</td><td>0.51</td><td>15</td><td>G10E15</td><td>0.74</td><td rowspan="3">Easy</td></tr>
<tr><td>16</td><td>G6E16</td><td>0.34</td><td>16</td><td>G10E16</td><td>0.71</td></tr>
<tr><td>17</td><td>G6E17</td><td>0.52</td><td>17</td><td>G10E17</td><td>0.75</td></tr>
<tr><td>18</td><td>G6E18</td><td>0.41</td><td>18</td><td>G10E18</td><td>0.61</td><td>Moderate</td></tr>
<tr><td>19</td><td>G6E19</td><td>0.42</td><td>19</td><td>G10E19</td><td>0.72</td><td>Easy</td></tr>
<tr><td>20</td><td>G6E20</td><td>0.57</td><td>20</td><td>G10E20</td><td>0.68</td><td rowspan="2">Moderate</td></tr>
<tr><td>21</td><td>G6E21</td><td>0.50</td><td>21</td><td>G10E21</td><td>0.68</td></tr>
<tr><td>22</td><td>G6E22</td><td>0.74</td><td>22</td><td>G10E22</td><td>0.80</td><td>Easy</td></tr>
<tr><td>23</td><td>G6E23</td><td>0.60</td><td colspan="4" rowspan="2"></td></tr>
<tr><td>24</td><td>G6E24</td><td>0.71</td></tr>
</table>

6.5.4.1.2. Discrimination index.

So as to calculate the discrimination index, the same dichotomous data to compute the difficulty index was used (0 = wrong answer; 1 = right answer). However, there was a need to distinguish between high- and low-achievers. To do so, the number of correct responses to the T/F and MC questions were placed in descending order and participants were classified into three balanced groups: A (high-achievers), B (medium-achievers) and C (low-achievers). In those cases when this was not possible because the total number of respondents was not divisible by three, the same number of participants had to be allocated to groups A and C, even if group B was left with fewer participants. Once the groups were created, the number of correct responses observed in each of them was noted down, and the difference between Groups A and C was calculated (i.e., the score of Group C was subtracted to that of Group A). Finally, to compute the discrimination index, the following formula was applied:

$$\text{Discrimination index} = \frac{\text{A - C}}{\frac{\text{N}}{3}}$$

where: A = *N* of correct responses in the high-achievers group; C = *N* of correct responses in the low-achievers group, and N = *N* of test takers.

The resulting value was interpreted according to the coefficients reported in Table 6.28:

Table 6.28
Discrimination index values according to del Rincón et al. (1995).

Value	Discrimination
≥ 0.40	Very good
0.30 – 0.39	Good
0.20 – 0.29	Moderate
≤ 0.19	Bad

As the results show (see Table 6.29), comprehension tests discriminated very well between skilled and less skilled respondents. In Grade 6, 91.66% of these tests had a very good discrimination index and only one was considered to be moderate. In Grade 10, 59% of the comprehension tests discriminated very well between high-achievers and low-achievers, and only 9% discriminated to a certain extent. At university, where tests had proved to be easier, the discrimination power remained high, since 50% of tests were good discriminators and 25% discriminated very well between good and bad test takers. Across levels, all tests discriminated between Groups A and C, proving that the comprehension tests designed for the intervention were reliable and they distinguished between high- and low-achievers. Besides, the results show that the three comprehension types -detailed, general and inferential- (Buck, 2001; Rodgers, 2013) are reliably measured, at least by the T/F and MC questions included in the first two exercises.

These two indices tell us that, overall, the content comprehension tests measured the construct in a global and comprehensive way, which allowed us to see how much learners had understood from the episodes. If the tests had been too easy or difficult, or had not discriminated well between learners, the results on content comprehension would be compromised.

Table 6.29
Discrimination index of the comprehension tests administered to Grade 6, Grade 10 and university participants.

Grade 6				Grade 10				University			
#	**Episode**	**Value**	**Discrimination**	**#**	**Episode**	**Value**	**Discrimination**	**#**	**Episode**	**Value**	**Discrimination**
1	**G6E01**	0.47	Very good	**1**	**G10E01**	0.30	Good	**1**	**G10E01**	0.37	Good
2	**G6E02**	0.41		**2**	**G10E02**	0.23	Moderate	**2**	**G10E02**	0.29	Moderate
3	**G6E03**	0.46		**3**	**G10E03**	0.33	Good	**3**	**G10E03**	0.32	Good
4	**G6E04**	0.34	Good	**4**	**G10E04**	0.49	Very good	**4**	**G10E04**	0.43	Very good
5	**G6E05**	0.47	Very good	**5**	**G10E05**	0.41		**5**	**G10E05**	0.37	Good
6	**G6E06**	0.42		**6**	**G10E06**	0.41		**6**	**G10E06**	0.25	Moderate
7	**G6E07**	0.40		**7**	**G10E07**	0.53		**7**	**G10E07**	0.45	Very good
8	**G6E08**	0.51		**8**	**G10E08**	0.36	Good	**8**	**G10E08**	0.30	Good
9	**G6E09**	0.29	Moderate	**9**	**G10E09**	0.41	Very good				
10	**G6E10**	0.51	Very good	**10**	**G10E10**	0.49					
11	**G6E11**	0.61		**11**	**G10E11**	0.37	Good				
12	**G6E12**	0.49		**12**	**G10E12**	0.45	Very good				
13	**G6E13**	0.59		**13**	**G10E13**	0.51					
14	**G6E14**	0.61		**14**	**G10E14**	0.48					
15	**G6E15**	0.54		**15**	**G10E15**	0.42					
16	**G6E16**	0.55		**16**	**G10E16**	0.38	Good				
17	**G6E17**	0.68		**17**	**G10E17**	0.39					
18	**G6E18**	0.53		**18**	**G10E18**	0.61	Very good				
19	**G6E19**	0.65		**19**	**G10E19**	0.48					
20	**G6E20**	0.58		**20**	**G10E20**	0.47					
21	**G6E21**	0.61		**21**	**G10E21**	0.22	Moderate				
22	**G6E22**	0.49		**22**	**G10E22**	0.39	Good				
23	**G6E23**	0.55									
24	**G6E24**	0.58									

6.5.5. Proficiency tests.

Participants in the thesis belong to different age groups and were at different stages of their FL learning development; therefore, measuring proficiency was necessary. Two common proficiency tests (a VS test and a dictation test) were administered to participants from all groups. The other proficiency tests were different, depending on the level of the participants: the tests administered in Grade 10 and university were the same (i.e., OPT), but they differed from those used in Grade 6 (i.e., Barcelona Age Factor -BAF- listening test), with lower-level learners.

As can be seen in Table 6.30, different types of tests were chosen. First, because listening skills were deemed crucial for the intervention, two listening tests were used: the listening part of the OPT (Allan, 2004) was taken by Grade 10 and university students and the BAF listening test was used in Grade 6 (Muñoz, 2006). Listening skills are also relevant for vocabulary acquisition, since research has shown that they are linked to VS (Miralpeix & Muñoz, 2018; Staehr, 2009) and that aural input is a good source for vocabulary acquisition (Elley, 1989; Maneshi, 2017; van Zeeland & Schmitt, 2013a). Apart from the listening tests, university students also took the grammar part of the OPT, to have a clearer picture of these learners' general proficiency. A dictation test was also taken by the three proficiency groups, as it can give an indication of segmentation ability: the better it is, the easier to identify word boundaries from oral input. The VS test (X_Lex and Y_Lex) was included as a comprehensive measure of general proficiency. Previous studies have revealed that VS is significantly linked to overall lexical proficiency, as well as to form and meaning recognition and recall (Montero Perez et al., 2014; Montero Perez et al., 2015; Peters et al., 2016), and that it plays a decisive role in the learning of vocabulary through video viewing (Peters & Webb, 2018; Suárez & Gesa, 2019).

It was also useful to check the adequacy of the TV series for each group of participants (e.g., lexical coverage). Further information on the proficiency tests is provided in the following sections.

Table 6.30
Summary of proficiency tests administered at each level.

Level	**Proficiency tests**					
	OPT Listening	OPT Grammar	Listening BAF	Dictation test	X_Lex (VS test)	Y_Lex (VS test)
Grade 6	✘	✘	✓	✓	✓	✘
Grade 10	✓	✘	✘	✓	✓	✓
University	✓	✓	✘	✓	✓	✓

6.5.5.1. Oxford Placement Test.

The OPT 1 (Allan, 2004) was used to assess listening and grammar. According to the manual, the OPT was calibrated with both NSs and EFL learners from different backgrounds and proficiencies to be able to grasp any differences in language knowledge. In the light of what the test designer states, the entire test was piloted with monolingual and multilingual learners in different language institutions for five years and with learners from 40 different nationalities, and the final test version was adapted on the basis of the results. The test is divided into two different parts: grammar (tapping into grammar, vocabulary and reading skills) and listening (focusing on reading and listening skills, and VS). Sample questions from both parts can be found in Appendix H.1.

The grammar part, which was only taken by the university group, consists of a written multiple-choice test that contains 100 sentences in which learners have to choose which of the three options given is the most suitable. When possible, these sentences are contextualized and linked to one another. The test is printed in four A4 pages, with 15, 35, 25 and 25 sentences respectively. For instance, two sample items of the test are:

Item #19 - Like any top sportsman Ali ***had to*** / *must* / *should* train very hard.

Item #96 - I think I'm expected to pick him up, ***aren't I*** / *don't I* / *are you*?

The listening part of the OPT, which was taken by both Grade 10 and university participants, contains 100 more items selected from a corpus of 'slips of the ear'; that is, situations when a non-native speaker's listening skills were inadequate and resulted in the transmission of the wrong message or in a breakdown in communication (Allan, 2004). In the test, participants are asked to listen to original sentences by NSs and decide which of the two options given corresponds to the actual words uttered in the recording. Both options are semantically and grammatically plausible, so test takers can only rely on their listening abilities to choose the most appropriate answer. During the test, learners can only listen to each sentence once. The listening test is printed in four A4 pages, the first one being introductory and the other three containing 35, 35 and 30 sentences, respectively. The audio lasts a total of 10 minutes and 33 seconds, with longer pauses when learners are required to turn the page. Two sample sentences of the test are:

Item #42 - We've gone through *today's* / ***two days'*** money in less than an hour.

Item #79 - We'll be letting them have a ***newer system*** / *new assistant* if they want one.

6.5.5.2. BAF listening test.

Primary school children took the BAF Project listening test. This test was originally designed, piloted and validated in one of GRAL's former projects (Muñoz, 2006). The test consists of 17 items, increasing in difficulty, each of them comprising an aural sentence and three possible pictorial representations. Learners' task is to listen one-by-one to the audio recorded by a NS and select the image that best represents each sentence. The audio recording begins by giving concise instructions to learners (i.e., "Choose the picture that corresponds with the sentence that you hear. You will only hear it once.") to continue with the first sentence. The test lasts 2 minutes and 58 seconds and it is printed on six A4 sheets of paper, with three questions in each of them, except for the last page, on which there are only two questions. Sample questions from the BAF Listening Test can be seen in Appendix H.2. One of the questions of the test is:

Item #11 - "Susan can see the garden when she is in bed."

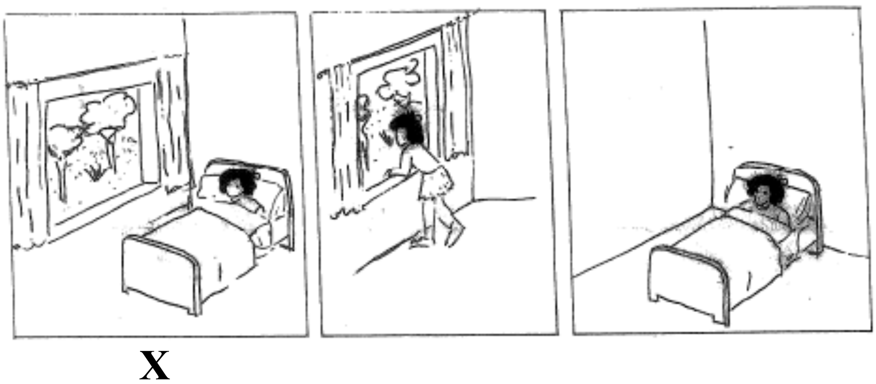


Figure 6.2 - Sample question of the BAF listening test.

6.5.5.3. Dictation test.

The dictation consists of a text of fifty words that learners are instructed to first listen to in one go. After that, it is broken down into shorter sets (each of them repeated twice) that learners have to write down. At the end, the whole text is read aloud one more time. This dictation was originally used in the BAF Project (see Muñoz, 2006), back in the late 1990s and early 2000s. However, and although the text of the dictation is exactly the same, it was recorded again in high quality by a NS originally from the United States. In total, the recording lasts 3 minutes and 9 seconds: after the text is read for the first time, there is a ten-second interval, each shorter set is followed by a six-second pause, and there are ten more seconds between the last set and the second reading. Learners are given an A4 sheet of paper with 12 lines (one per sentence) on which they have to write the dictation. The text that learners have to write is the following (// indicates when there is a pause in the audio):

"the first week in Jamaica // was wonderful // the sun shone every day // and the sea // was as blue // as in any travel poster // the girls spent the days // on the little beach // below the house // there was no road // from the house to the beach // just a winding narrow path"

6.5.5.4. Vocabulary size test: X_Lex and Y_Lex.

From the different VS tests available, X_Lex v2.05 (Meara, 2005; Meara & Milton, 2003) and Y_Lex v2.05 (Meara & Miralpeix, 2006) were chosen to measure VS because they are computerised, quick to administer and have been shown to work in the same context at different proficiency levels (Miralpeix, 2012). X_Lex analyses vocabulary included in the first five thousand words whereas Y_Lex taps into vocabulary included in the 6k-10k word range. On these computer-based tests, learners' task consists in selecting a happy face when they know the meaning of the word shown on the screen or the sad face when they are not sure, or they do not know it. The tests contain a number of pseudo-words in order to control for guessing (i.e., learners' scores are adjusted downwards if they claim to know one of these words).

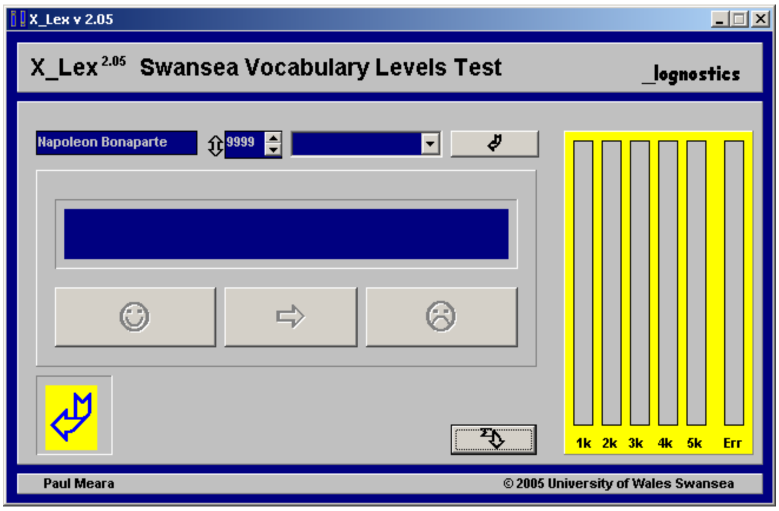


Figure 6.3 - Screen capture of how X_Lex v2.05 looks like.

6.6. Procedure

6.6.1. Pilot study.

The pedagogical intervention was conducted during the 2015-2016 academic year in Grades 6 and 10 and during the first academic term of the same school year at university. Previously, a pilot study was conducted in May 2015 to try out some of the materials specially designed for this intervention and the procedure to be followed in the main study. The pilot study took place in the same institutions that the following year would collaborate in the intervention. Two of the groups in which materials were piloted were Grade 6 and first-year university students, the same level as the groups that would participate the next academic year, while the other was a Grade 11 group (not Grade 10). This is because the intervention was initially planned to be conducted in Grade 11; the reasons being that Grade 11 is the first year of post-compulsory education and students who decide to continue studying tend to be the more committed and willing to participate in this type of studies. Nevertheless, after conducting the pilot study and before the beginning of the intervention, the school board decided that Grade 10 would be the most appropriate population for the experiment. They argued that too many curricular activities were already planned in Grade 11 and there was no time left for the intervention. That is why the pilot study was conducted with Grade 11 EFL learners, whereas the main study was performed in Grade 10. Actually, the difference in proficiency between these two levels is small so it does not compromise the results from the pilot study.

The pilot study consisted of a single viewing session in which the TV series and some of the materials were tested. For logistic and temporal reasons, the entire pedagogical intervention could not be piloted as this would have implied an extra year of data collection. However, all students participating in the pilot watched the TV series since we were especially interested in evaluating the materials of the EG and the organization of a session with TV viewing.

For the pilot, an episode from each TV series was selected and materials were prepared for each of them: a short vocabulary pre-test, vocabulary pre-task, while-watching task, vocabulary post-task and comprehension test (see Appendix I). All of them have already been introduced in the 'Instruments' section, but before having the final versions presented, some adaptations were made after the pilot:

a) In the vocabulary pre- / post-tests and the post-tasks, two repetitions of each TW, instead of three, were finally included in the audio files because learners commented that two were enough to write down the TWs' forms. Including three repetitions of all 40 TWs would have made the test longer.

b) The introductory exercise in the vocabulary pre-tasks, aimed at introducing the plot of the episode learners were about to watch, was eliminated. This first exercise took too long and did not fulfil the objective as some students were not paying much attention to answer the questions.

c) L2 subtitles for Grade 6 learners were discarded, since participants themselves claimed that they were too difficult to follow, and they did not help them to understand the storyline of the episode. After consultation with the teachers and other researchers, it was decided to use L1 subtitles throughout the academic year. As research has already proven (Danan, 2004), L1 subtitles are especially beneficial for low-level learners, so sixth-graders needed the extra help of their L1 to succeed in the intervention.

d) Because of c), and although it was well received by primary school learners, *Boy Meets World* (Jacobs et al., 1993), one of the TV series used in the pilot study in Grade 6, was replaced by *Wizards of Waverly Place* since the former did not have Spanish subtitles available.

e) The open-ended question in the while-watching task asking participants to provide a succinct summary of what had happened in the first half of the episode was substituted for two MC questions. As this task only aimed at keeping the participants' attention on the video, and no measure of vocabulary acquisition or content comprehension was calculated on its basis, it took too long to complete: learners wrote extended summaries of the first half of the episode.

f) Instead of providing Grade 6 learners with the TWs' forms in the vocabulary post-task, it was decided to change its format and that the written form would not be given. Indeed, it was seen that participants did not pay attention to the audio file as they did not need to resort to it to identify which of the TWs was being asked.

g) The number of TWs selected in each episode was reduced from eight to five to shorten the duration of the pre- and post-tests and because some learners expressed that they could not cope with so many words in such a short period of time.

Furthermore, as part of the pilot, a questionnaire (see Appendix I.6.) was added at the end of the session to collect students' opinions about the TV series and how they would feel about participating in the intervention. It included questions like:

i. Did you like the episode?
ii. Would you like to watch more episodes of the same TV series?
iii. Have you understood what was going on in the episode?
iv. Did you read the subtitles?
v. Did they help you to understand the episode?
vi. Would you like to watch the episode again with English subtitles? (only for Grade 6 learners who had seen the TV series with L1 subtitles)
vii. Do you think you have learned new vocabulary?

Results showed that vocabulary acquisition took place between the pre-test and the vocabulary post-task. However, as these two exercises had different formats, gains could not only be attributed to the experience of having watched a TV series. Results indicated that vocabulary acquisition from exposure to television viewing was possible and answers to question vii) in the questionnaire also showed that learners had the feeling of having learned new vocabulary, with the vast majority of learners at all levels reporting that they learned a lot of or some vocabulary during the viewing session.

The TV series were also well-received. All Grade 6 EFL learners claimed that they liked *The Suite Life of Zack and Cody*; none of the 25 participants reported that they disliked the TV series and 68% said that they liked it very much. 42.86% in Grade 11 and 83.33% at university reported that they liked *Seinfeld*. This TV series was chosen to be piloted because *I Love Lucy* had already been piloted with a university group (Cokely & Muñoz, 2019) and the results showed students' positive attitudes towards it.

All the data gathered during the pilot study was useful because, in the next three months, before the pedagogical intervention, the problematic aspects found in the materials and the design could be solved, and all the necessary changes made before the main study started.

6.6.2. Pedagogical intervention.

The pedagogical intervention took place between September 2015 and June 2016, with the vocabulary delayed test being administered in February 2017 (check Table 6.31 to see the calendar of the pedagogical intervention). For logistic reasons and in order to compute vocabulary gains after sustained exposure to TV series, the study was divided into three terms: Term 1 (September - December), Term 2 (January - March), and Term 3 (April - June). Prior to the beginning of the academic year, a meeting was held with all the teachers involved in the intervention. In this meeting, the design of the intervention and the characteristics of all the tests and tasks were explained, and explicit instructions were given. Besides, all the questions were answered and the researcher's contact phone number and email were also provided in case there were some other questions. Teachers were also told that the researcher would be in charge of the first viewing session and all the vocabulary pre- and post-tests. Likewise, all the proficiency tests and other tasks that required the use of the computer lab were led by the researcher as well. The rest of the viewing sessions were teacher-led, once the procedure was clear and the normal functioning of the class was guaranteed.

At the beginning of the academic year, proficiency tests (i.e., listening part of the OPT -Grade 10 and university-, grammar part of the OPT -university-, BAF listening test -Grade 6- and dictation test -all proficiency groups-) were administered to serve as entry-level measures. Likewise, the VS tests were also administered towards the beginning of the academic year, together with other cognitive and aptitude tests that are not part of the dissertation. The VS tests were taken individually using laptops, with the presence of the researcher and learners were given feedback on their results at the end of the testing session.

In each of the terms, the same procedure was followed: at the beginning of the term, participants took the vocabulary pre-test, divided in the two parts already explained in the instruments section. The audio was played in the overhead speakers in the classrooms, and time was given to go through the test before beginning. The two parts were printed on two separate A4 pages which were distributed one at a time. Between the parts, there was a short pause and students were given some time to complete their answers. In the following weeks, viewing sessions took place (Figure 6.4 presents an outline of what was done each term).

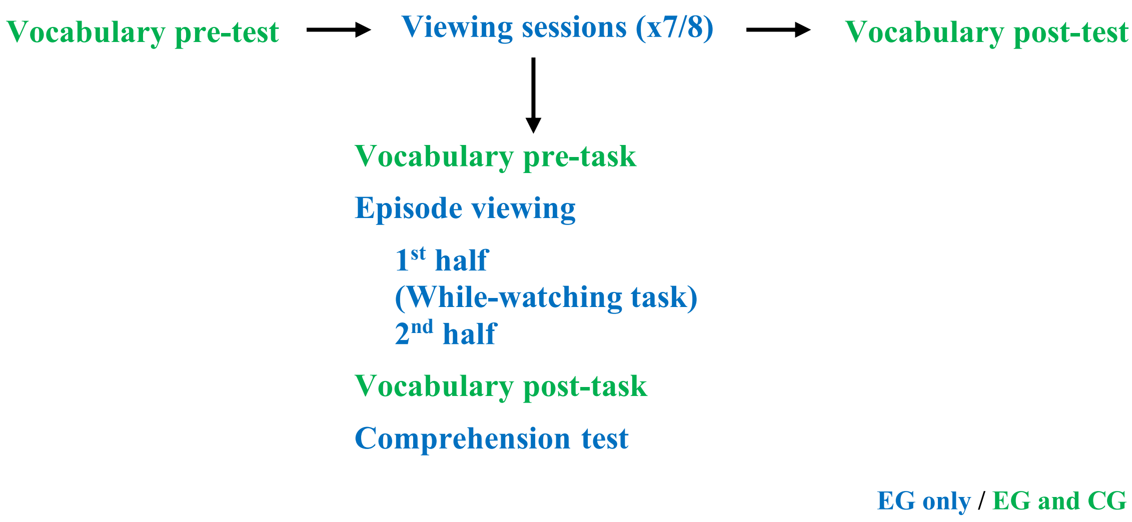


Figure 6.4 - Summary of the procedure followed each term of the intervention.

In each of these sessions, and for the whole academic year, EGs started by doing the vocabulary pre-task. The pre-task was always done individually or in small groups and it usually lasted no more than five or ten minutes. It was presented on a one-sided sheet of paper and it was corrected immediately afterwards. This way, doubts and questions learners had regarding the TWs could be answered. The teacher or researcher collected the vocabulary pre-tasks after their correction, so that students could not resort to them when completing other activities. Afterwards, participants watched the L1 or L2 subtitled episode, which was projected onto the classroom's whiteboard using an overhead projector. The image was big enough so that students could see it and read the L1 or L2 subtitles. Participants were explicitly told to sit near the whiteboard so as to avoid any viewing and reading problems.

The while-watching task was carried out in the middle of the episode. The teacher / researcher stopped the video at the indicated time and distributed the tasks, always printed on an A5 page. Participants were given about two minutes to answer the two MC questions. Immediately afterwards, it was corrected in class (it did not aim at measuring content comprehension and was not related to the comprehension questions in the tests at the end of the session). It should be remembered that the aim of this task was to keep the participants' attention on the videos, which lasted no more than 25 minutes. After some sessions, participants and teachers pointed out that this task was fairly disruptive, and it did not meet the aim it was designed for because learners were all paying attention and enjoying the viewing sessions. For these reasons, during the second term, some of the teachers involved in the intervention decided to read the questions aloud, asking them to one or two students in class. This task was finally eliminated at the beginning of Term 3. However, with longer episodes lasting about an hour (like thrillers and drama series), it may be reasonable, and even advisable, to include this task to keep participants focused. After the while-watching task, participants in the EGs continued watching the second half of the episode.

At the end of the viewing session, learners in the EGs completed the vocabulary post-task. It was distributed on a one-sided A4 page by the teacher and he / she gave one or two minutes to students to go over it. Afterwards, the audio file was played using the overhead speakers, and learners were given some more time to complete their answers. In contrast to the vocabulary pre-task, the post-tasks were not corrected in class. The researcher was in charge of correcting them in order to check whether immediate recalling of the TWs was easy or difficult for the learners. Besides, it could be seen whether students understood and remembered the TWs and (dis)confirm that immediate recalling of the target vocabulary was high. The post-task was a way of providing learners with some more exposure to the TWs, as purely incidental learning from video exposure would have been too challenging (e.g., Pujadas, forthcoming). If students had been given the opportunity to correct the post-task, it is highly probable that they could have changed their answers or marked one question as right when it was wrong. As expected, the scores on the post-tasks were very high regardless of the experimental condition, with many participants always reaching the maximum score. For this reason, the scores of the vocabulary post-tasks are not reported in this doctoral dissertation; besides, assessing immediate recall was not the aim of the thesis.

Once the EGs completed the vocabulary post-task, the comprehension test was distributed. It was never given to students until they had not handed in the vocabulary post-task, just in case the latter hinted at some of the answers of the former. The comprehension test was printed on one side of an A4 page, and it took learners about five to seven minutes to go through and complete it. They were encouraged to answer all the questions and were explicitly told that wrong answer did not penalize. The comprehension test was not corrected in class, as learners could have changed their answers while marking, which would have affected the results of the thesis.

During the viewing sessions, the CGs followed exactly the same procedure as the EGs, the only difference being that, while the EGs were watching the episode, teachers from the CGs were instructed to do other activities to fill in class time. No explicit activities were designed for the CG, although some were not allowed. For instance, no TV or movie viewing was accepted, and teachers were instructed to design activities that did not involve too much listening and no contact with the TWs. As they had the list of the TWs from all the viewing sessions, care was taken to avoid any contact with them during the CGs' activities. The CGs continued with their regular syllabus instead of watching the videos and, of course, they did not take the comprehension test as they had not seen the episode.

After the seven / eight viewing sessions of each term, the vocabulary post-test took place. The post-tests were the same as the pre-tests. However, in those terms when seven episodes were watched (instead of eight) (Terms 2 and 3 in Grade 10), the post-test was shortened to only include the TWs which had been actually introduced in class.

Data collection finished in February 2017, when the vocabulary delayed test was administered to students that took part in the intervention in Grade 6 (now in Grade 7) and Grade 10 (now in Grade 11). The same materials as in Term 3 were used (with the minor changes explained in section 6.5.2.3.) and the same procedure at that time was followed. Table 6.31 shows the calendar of the entire pedagogical intervention; differences between Grade 6 and 10 are also indicated:

Table 6.31
Calendar of the pedagogical intervention.

Term	Week	EGs	CGs
Term 1 (September - December 2015)	Week 1	Proficiency tests and vocabulary pre-test T1	
	Week 2	Viewing session 1	Session 1
		VS tests	
	Week 3	Viewing session 2	Session 2
	Week 4	Viewing session 3	Session 3
	Week 5	Viewing session 4	Session 4
	Week 6	Viewing session 5	Session 5
	Week 7	Viewing session 6	Session 6
	Week 8	Viewing session 7	Session 7
	Week 9	Viewing session 8	Session 8
	Week 10	Vocabulary post-test T1	
Term 2 (January - March 2016)	Week 11	Vocabulary pre-test T2	
	Week 12	Viewing session 1	Session 1
	Week 13	Viewing session 2	Session 2
	Week 14	Viewing session 3	Session 3
	Week 15	Viewing session 4	Session 4
	Week 16	Viewing session 5	Session 5
	Week 17	Viewing session 6	Session 6
	Week 18	Viewing session 7	Session 7
	Week 19	Viewing session 8 -Grade 6-	Session 8 -Grade 6-
		Vocabulary post-test T2 -Grade 10-	
	Week 20	Vocabulary post-test T2 -Grade 6-	
Term 3 (April - June 2016)	Week 21	Vocabulary pre-test T3	
	Week 22	Viewing session 1	Session 1
	Week 23	Viewing session 2	Session 2
	Week 24	Viewing session 3	Session 3
	Week 25	Viewing session 4	Session 4
	Week 26	Viewing session 5	Session 5
	Week 27	Viewing session 6	Session 6
	Week 28	Viewing session 7	Session 7
	Week 29	Viewing session 8 -Grade 6-	Session 8 -Grade 6-
		Vocabulary post-test T3 -Grade 10-	
	Week 30	Vocabulary post-test T3 -Grade 6-	
February 2017	Week 31	Vocabulary delayed test -Grades 6 and 10-	

6.7. Test scoring and data analysis

6.7.1. Scoring criteria.

All tasks and tests that were not corrected in class (i.e., vocabulary pre- and post-tests, vocabulary post-task, and content comprehension tests) as well as the proficiency tests were corrected by the researcher following the criteria explained below (see Table 6.33 for the maximum scores learners could get in each test).

6.7.1.1. Proficiency tests.

Proficiency tests were corrected following standardized conventions. Regarding the OPT, the guidelines by Allan (2004) were used. Accordingly, one point was awarded for each correct answer and no points were given for incorrect or unanswered questions, wrong answers were not penalised. Hence, the maximum score learners could obtain on the test was 200 points: 100 for the grammar part (university students only) and 100 more for the listening part (university and secondary school learners).

The BAF listening test was also scored dichotomously: 1 point was given to each correct answer and 0 to wrong answers, so the maximum score learners could get was 17 points, one for each item. As far as the dictation is concerned, the maximum score was 50. However, two criteria were adopted to account for partial knowledge of the fifty words presented in the recording:

1. In the first, more strict approach, no orthographical mistakes or deviant forms of the words were allowed. For instance, if the word 'winding' was written as 'wynding', no points were assigned.
2. In the second, more lenient criterion, partially correct words were taken into account, so simple spelling errors (not more than one per word) were marked as partially correct. For instance, the above-mentioned spelling of 'wynding' was given half marks. So were words that showed some degree of recognition of the target items; for example, the use of inflected forms or homophones. To name some, 'houses' (instead of 'house') or '1st' (instead of 'first') were given some credit. Finally, words containing two or more errors were labelled as wrong (e.g., 'tavell' instead of 'travel' or 'rout' as opposed to 'road'). Segmentation problems were also accounted for and no points were assigned.

It was seen that the task was not too challenging for participants (for instance, always following the strict criterion, sixth graders obtained a mean score of 23 words, Grade 10 learners 42 words, and the university group 43 words), and that the two scoring criteria were positively and significantly correlated (r=.990, p=.000). Therefore, scores from the first strict criterion were used in the analysis.

The VS tests were automatically scored by the software itself. However, both X_Lex and Y_Lex provide two scores: the raw score of those words that learners claim to know, and the corrected or adjusted score, controlling for the number of pseudo-words that learners said they were familiar with. Following the tests' manual, the adjusted score was adopted. Participants who claimed to know six or more pseudo-words were eliminated from the sample, as their scores were not reliable (Miralpeix, 2012). The total VS scores reported in the dissertation was the addition of the scores on both tests (X_Lex and Y_Lex) -in Grade 10 and university students- or the score on the X_Lex test only -in Grade 6 learners-.

6.7.1.2. Pre-, post- and delayed vocabulary tests.

Pre-, post- and delayed vocabulary tests were scored following three different criteria: the first accounted for full knowledge of the target items whereas the remaining others took partial knowledge into account. Assessing for partial knowledge was deemed necessary in response to feedback received by the scientific community (Ellis, personal interview; Rodgers, personal interview).

Following the full knowledge criterion, much stricter than the other two, no spelling errors were accepted (e.g., 'miten' instead of 'mitten' was considered incorrect). As Webb (2007, p. 55) argues:

> This was because the learners were given the phonological forms of the target words as a cue to recall. Since the participants were at the intermediate level and were likely to have learned most if not all of the rules of spelling, phonological cues would be enough to at least lead them to write a close approximation of the target words. If responses with minor spelling mistakes were marked as correct, then it could not be determined whether it was due to repetition -an encounter with the target words in the tasks- or the phonological prompt.

In the case of verbs, which had been dictated as 'to + infinitive' forms, only the bare infinitive was required. Hence, the first measure derived from the vocabulary tests was the number of TWs' forms which students got right (a maximum of 35 or 40).

As far as the TWs' meanings were concerned, learners were instructed to give an L1 translation of the TWs, although L1 / L2 synonyms and definitions in any of the three languages (English, Catalan or Spanish) were also accepted. Besides, participants were instructed to include all the TWs' meanings they were aware of. In those cases where the TW had multiple meanings (around 6.5% of the TWs in Grade 6 and 10% in Grade 10 and university were considered polysemous), only the meaning shown in the pre-task, the post-task and the episode was accepted as a correct response. Although in the pre-tests learners did not know which meaning was going to be illustrated in the episodes, the researcher only marked as correct those that reflected the meanings about to be shown. However, on the post-test and the delayed test, students were warned about this in advanced. For example, according to the Oxford English Dictionary, the meaning of 'date' which was shown in the TV series was 'a social or romantic appointment or engagement', so translations like 'fecha' or 'data' and definitions such as 'the day of the month or year as specified by a number' (date, n.d.) were not considered correct.

Regarding the two partial knowledge criteria, they differed in their degree of strictness: while one was considered more sensitive, the other was tougher. However, none of them was as strict as that already explained in the previous paragraphs.

According to the tough criterion, no mistakes were allowed in one-syllable words so, for instance, 'to wipe' was marked as wrong if written 'whipe'. Moreover, in polysyllabic words, one spelling mistake was tolerated as long as it did not affect the pronunciation of the TW, always bearing in mind the different ways in which graphemes can be pronounced in English and the mother tongue of the participants (namely, Spanish and Catalan). For example, 'blirb' was accepted as an alternative to 'blurb' (and given half marks), but not 'blorb'. Both 'i' and 'u' are graphemes that represent the phoneme /ɜː/, but 'o' is not a representation of such phoneme in English. Another example could be 'cavin' (for 'cabin'), which was marked as partially correct because 'v' and 'b' are representations of the phoneme /b/ in Spanish and central Catalan, the variety spoken in Barcelona, Mataró and Mollet del Vallès, where data collection took place. Finally, words that are pronounced the same but have different meanings were classified as incorrect ('brake', for 'to break').

The other scoring criterion accounting for partial knowledge, a bit more sensitive, only differed from the previous one in the sense that one mistake was allowed in monosyllabic words, provided that it did not alter pronunciation. To continue with the same example, 'whipe' (for 'to wipe') was given some credit (half a point). Finally, it was not necessary to conduct a partial knowledge assessment for TWs' meanings, since several options were already accepted from the very beginning.

Once all the pre- and post-tests were scored following the three criteria explained before (full knowledge, tough partial knowledge and sensitive partial knowledge), statistical analyses showed that there were very high significant correlations between them. To work as an example, this is the data for Term 2 for both proficiency groups:

Table 6.32
Two-tailed Pearson *r* correlations between full and partial knowledge criteria at T2 in Grades 6 and 10.

Level	Testing time	Two-tailed correlations	Tough partial knowledge criterion	Sensitive partial knowledge criterion
Grade 6	Pre-test	Full knowledge criterion	*r*=.954	*r*=.923
			p=.000	
	Post-test		*r*=.973	*r*=.957
			p=.000	
Grade 10	Pre-test		*r*=.995	*r*=.990
			p=.000	
	Post-test		*r*=.988	*r*=.985
			p=.000	

As the three criteria were significantly correlated and all the correlations were very strong, it was assumed that the scoring criterion would not affect the results. Hence, it was not deemed necessary to report all the scores using the three methods, and only the full knowledge criterion will be referred to throughout the dissertation.

Finally, as far as the delayed vocabulary test is concerned, it should be noted that the TWs participants had not had the opportunity to learn during T3 because they had missed one viewing session were excluded from the analysis. Once done, only those TWs that had been learned during the intervention at T3, that is, those that were unknown on the pre-test T3 but known on the post-test T3 for form and meaning, were later classified as ‘maintained’ or ‘lost’ on the delayed test. By ‘maintained’, we mean that the TW was again answered correctly eight months after the end of the intervention, whereas words which were not answered correctly on the delayed test were labelled as ‘lost’. This gave a clear picture of the percentage of words that were still remembered after considerable time without planned exposure to captioned TV series and the TWs, as it discarded those items that were answered correctly on both the preand post-test in Term 3 and 'unlearned' words, that is, those words that were answered correctly on the pre-test but incorrectly on the post-test.

6.7.1.3. Comprehension test.

The comprehension tests were scored dichotomously. One point was given for each T/F and MC question that was answered correctly, and no points were subtracted for wrong responses. In these two exercises, a total of 10 points could be obtained (five for the T/F questions and five for the MC questions). In the last exercise in which participants had to order eight statements chronologically, one point was awarded for each sentence placed in the right position, with no deductions applied for unfitting responses. This excluded sentences #1 and #5, which were already given by the researcher to avoid a null score. Hence, the maximum score that learners could obtain in this last exercise was six points. This implies that the highest score for the comprehension test was 16 points (5+5+6).

Table 6.33
Maximum scores learners could get in each of the tests used in the pedagogical intervention.

Test	Maximum score
OPT listening test	100 points
OPT grammar test	
Listening BAF	17 points
Dictation	50 points
X_Lex	5,000 points
Y_Lex	
Vocabulary pre-test	35 / 40 points
Vocabulary post-test	
Vocabulary delayed test	
Comprehension test	16 points

6.7.2. Absolute and relative lexical gains.

As the pre-, post- and delayed vocabulary tests were identical, lexical gains for each of the terms were calculated (Nation & Webb, 2011). The aim of calculating gains was to observe students' vocabulary learning throughout the year. Firstly, it should be noted that two types of gains (absolute and relative) were calculated for three different variables (form, meaning and form + meaning altogether). Separate calculations were done as the form and the meaning of words are two different aspects of lexical knowledge (Nation, 2013). The form-meaning link was assessed because associating a word form with its meaning is an initial and simple form of lexical knowledge.

Regardless of the type of gains calculated, the raw scores of the pre- and post-tests for the three variables (i.e., the number of TWs answered correctly) were converted into percentages because some of the participants had attended all viewing sessions whereas others had missed one. Previously, the scores of those that had only attended *n-1* sessions (i.e., six or seven, depending on the term) had been adjusted to exclude the five TWs they had not had the opportunity to learn. Thus, the number of TWs on which participants were tested differed on an individual basis, sometimes being 30, and others 35 or 40.

In order to calculate absolute gains, the score on the pre-test was subtracted to the score of the post-test:

Absolute gains = Post-test score - Pre-test score

Because gains were very much dependent on the number of words learners knew at the beginning of the term and on the room for improvement each participant had, another measure controlling for this difference and taking into consideration the varying opportunities for learning was needed. Relative gains are considered a more fine-grained measure of lexical growth in experimental conditions (Horst et al., 1998) and they exhaustively control for learners' previous knowledge of the items tested. They were first defined by Shefelbine (1990) and later used in different vocabulary studies (e.g., Peters & Webb, 2018; Rodgers, 2013). Relative gains for participants were calculated for each of the three variables (form, meaning and form + meaning) using the following formula:

$$\text{Relative gains for participants} = \frac{N \text{ of TWs learned}}{N \text{ of items tested - } N \text{ of TWs known}} \text{ x } 100$$

where: 'TWs learned' = N of items that were answered incorrectly on the pre-test and correctly on the post-test; and 'TWs known' = N of items which were answered correctly on both the pre- and the post-test. The 'N of items tested' was always 30, 35 or 40, depending on the participant.

6.7.3. Statistical analyses.

To fulfil the aims of the doctoral dissertation and give an answer to the RQs, several statistical analyses were performed.

6.7.3.1. RQ1: TV viewing and L2 vocabulary learning.

To answer the first RQ ('Does extensive viewing of subtitled TV series enhance L2 vocabulary learning?'), relative gains for form and relative gains for meaning were used. As the number of participants in the final sample varied across terms, the analyses were conducted term-independently, and gains were calculated for T1, T2 and T3.

First of all, mixed between-within ANOVAs, with time and condition as factors, were run to investigate the evolution of relative gains throughout the year. At university, though, this statistical analysis was not run, since there was just one academic term. Then, to analyse if improvement occurred between the beginning and the end of the term, a series of paired-samples *t*-tests or Wilcoxon signed-rank tests (depending on the normality of the data) were run with the scores on the pre- and the post-tests as dependent variables. Finally, and in order to know whether exposure to video viewing was beneficial for the participants' vocabulary development, the relative gains obtained by the two experimental conditions (EG vs. CG) were compared term independently. Therefore, independent samples *t*-tests or Mann-Whitney *U* tests were conducted to know if such differences were statistically significant.

6.7.3.1.1. RQ1.1.: TV viewing and L2 vocabulary retention.

The aim of RQ1.1. ('To what extent does sustained exposure to subtitled TV series affect vocabulary retention?') was to see whether knowledge gained during the intervention was maintained or forgotten months after its end, and whether modality of input played a significant role in such process. To answer it, the percentage of word forms and word meanings 'maintained' (see section 6.7.1.2.) on the delayed test was compared across experimental conditions. In order to do so, independent-samples t-tests or Mann-Whitney U tests were run depending on the normality of the data analysed.

6.7.3.2. RQ2: TV viewing and content comprehension.

To answer the second RQ, ('Does content comprehension of subtitled TV series change across successive episodes viewed?'), the scores of the comprehension tests were taken into account, after having converted them into percentage scores. Repeated Measures (RM) ANOVAs with comprehension scores as the dependent variables and time as factor were conducted for each of the three proficiency groups. However, as the number of participants was fairly reduced, there were not enough cases to conduct a RM ANOVA with 22 or 24 testing times, that is why a RM ANOVA was run for each academic term (7 or 8 testing times) and proficiency group, instead of the whole academic year.

6.7.3.3. RQ3: TV viewing and learners' proficiency level.

In order to evaluate how L2 vocabulary learning and content comprehension are affected (or not) by learners' proficiency level, just the data from the EGs (and not from the CGs) were taken into consideration, because they were the only students exposed to TV series. Vocabulary gains (form and meaning) and mean content comprehension scores of Grade 6, Grade 10 and university learners were compared term-independently. In Term 1, one-way between-groups ANOVAs or Friedman tests were conducted with grade as factor and scores as dependent variables. In the other two terms, due to the absence of the university group, independent samples t-tests or Mann-Whitney U tests were run with the data from Grade 6 and Grade 10 learners.

CHAPTER SEVEN - RESULTS

Throughout the next subsections, results will be presented following the three RQs proposed: on (1) vocabulary learning and retention, (2) content comprehension and (3) the mediating role of proficiency in the learning and development of these two aspects. For the first two RQs, results are presented for each proficiency group that the dissertation analyses. Hence, in each RQ, the results from the university group will be presented first, since most research so far has been conducted with undergraduate students of similar characteristics and the results of the present study will build on what has been observed in the past. Next, the results from Grade 10 will be outlined since these participants were also exposed to L2 subtitled materials and, during Term 1, they watched the same episodes and did the same tasks as the university participants. Then, the results from Grade 6 learners will be described, as this group watched different TV series subtitled in L1. Therefore, the presentation of the results follows a descending order of proficiency, from more to less advanced, which will help in their interpretation afterwards. In the results of RQ3, at the end of this chapter, the three EGs will be compared.

In all the analyses presented in the next sections for each RQ, the same procedure was followed: before conducting any inferential tests, data was thoroughly screened, and outliers were inspected and removed from the dataset if necessary. In addition, assumptions for each analysis were also checked (e.g., normality, homogeneity of variances, sphericity, homogeneity between the intercorrelations), and decisions were taken according to what was observed (e.g., non-parametric tests were selected when data failed to reach a normal distribution).

7.1. Proficiency level of university, Grade 10 and Grade 6 learners

First of all, the scores on the proficiency tests were examined so as to make sure that the three groups actually differed in proficiency and could be treated separately, and that no proficiency differences existed between EGs and CGs (see also section 6.4.).

To do so, a compound measure of the scores on the different proficiency tests was calculated. In Grade 6, this included the BAF listening test, the dictation test and the X_Lex VS test. For Grade 10 and university students, it included their scores on the dictation test, the OPT listening test and their VS score (calculated on the basis of X_Lex and Y_Lex tests). Participants' proficiency scores were converted into percentages and the average of the three tests was calculated, and this was statistically compared between levels and experimental conditions. Data was submitted to a one-way between-groups ANOVA once it was checked that normality assumptions were not violated. Results revealed significant differences between groups ($F(2, 125) = 41.767$, $p = .000$) and Tukey HSD post-hoc tests showed the sixth graders' scores on the proficiency tests ($M = 49.77$; $SD = 11.43$) were significantly lower than tenth graders' ($M = 62.80$; $SD = 6.77$) ($p = .000$) and university students' ($M = 68.69$; $SD = 8.36$) ($p = .000$). Besides, the scores obtained by Grade 10 students were significantly lower than those obtained by the university group ($p = .006$).

The scores on this compound proficiency measure were also compared across the EG and the CG at each level. No significant differences in proficiency were found between the two conditions at university (EG - $M = 68.95$, $SD = 8.39$; CG - $M = 68.23$, $SD = 8.58$) ($t(42) = .271$, $p = .788$). At Grade 10, differences were also non-existent (EG - $M = 63.06$, $SD = 7.70$; CG - $M = 62.49$, $SD = 8.27$) ($t(49) = .255$, $p = .799$). Finally, the same was found in Grade 6 ($M =$

53.34, *SD* = 10.05; CG - *M* = 46.80, *SD* = 11.92) ($t(31) = 1.682$, $p = .103$). Therefore, the two experimental conditions at each of the three levels were comparable in terms of proficiency, ruling out the possibility that this factor could explain any differences between conditions.

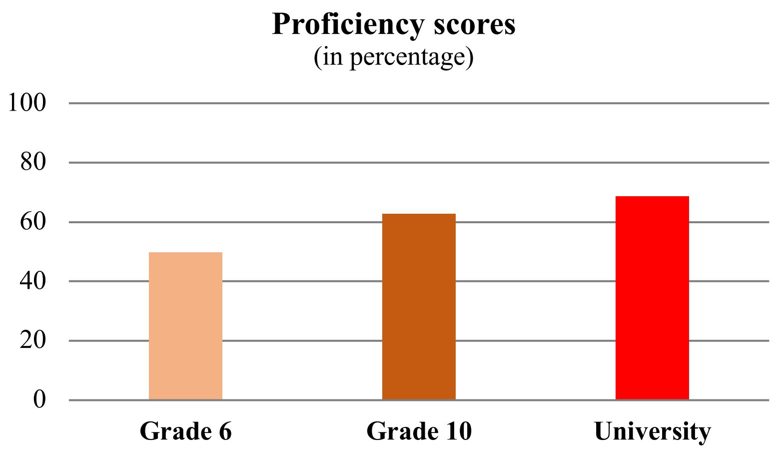


Figure 7.1 - Compound proficiency scores according to proficiency groups.

7.2. RQ1 - Vocabulary learning from subtitled TV series

The first RQ tapped into vocabulary learning from subtitled TV series and traced participants' vocabulary development during one academic term (in the case of university learners) or one academic year (in the other two proficiency groups). As stated in Chapter 6, vocabulary gains were calculated using the relative gains formula. This was applied to form, meaning, and form + meaning of the TWs. However, after seeing the results for the three variables, it was decided to only report the scores for 'form' and 'meaning', not 'form + meaning'. This decision was adopted because the scores for 'meaning' and 'form + meaning' were nearly the same and presenting both results could have been redundant. When participants knew the meaning of a word, they also knew the form. Put differently, there were very few cases in which learners knew the meaning of a TW and not its form and these basically were due to serious orthographic mistakes (e.g., the TW 'fungus' was written as 'fungest' even though the translation was correct, 'fongs / hongos', or the TW 'whistle' was translated as 'xiulet / silbato' although it was spelt 'wissel'). Actually, the mean difference between the relative gains for 'meaning' and 'form + meaning' throughout the year and across proficiency groups was:

- University → 4.55% in the EG and 4.58% in the CG
- Grade 10 → 4.10% in the EG and 2.90% in the CG
- Grade 6 → 2.82% in the EG and 3.94% in the CG

Based on this, it must be borne in mind that, when relative gains for 'meaning' are reported, they imply that learners were also familiar with the TW forms in most cases.

In the sections that follow, the descriptive results from the pre- and post-tests as well as the relative gains will be presented first, followed by the inferential statistics that give an answer to the RQ. This procedure will be repeated for the three proficiency groups (university, Grade 10 and Grade 6). For Grade 10 and Grade 6, who followed the intervention throughout the academic year, a RM ANOVA for the whole year will be presented before looking closer at each term. It should be reminded that some of the participants did not attend one of the sessions of the intervention in a given term, so that is why inferential tests were run with the percentage scores. However, both raw and percentage scores are reported in this chapter to have a clear idea of how many TWs participants knew.

By looking at the scores of all groups, we see that, at the descriptive level, there is a tendency for the EGs to score higher than the CGs on the post-tests and EGs' relative gains also seem to be larger although sometimes differences are very small. We can also see that the scores on the pre-tests are lower than those obtained on the post-tests. High variability is observed across all measures, as shown by the high standard deviations and the difference between minimum and maximum scores. For example, in each term, there were participants that did not learn new vocabulary whereas others learned up to 50% of the TWs. It is also evident from our findings that the scores for form are always higher than the scores for meaning learning in all conditions, proficiency groups and tests.

7.2.1. Vocabulary learning at university.

As far as university students are concerned, they participated in the intervention during one academic term and watched the episodes with L2 subtitles. In terms of the number of participants, the final sample included 61 learners in total: 38 belonging to the EG and 23 to the CG. The pre- and post-test's descriptive results (means, standard deviations, minimum and maximum scores) of both groups are found in Table 7.1. A set of paired samples *t*-tests and Wilcoxon signed-rank tests (see Table 7.1 as well) corroborated that significant progress was made between the pre- and the post-test (*p*s = .000 in all cases) by both experimental conditions.

Table 7.1
Pre- and post-test's results for university students.

	EG (*N*=38)								**CG** (*N*=23)							
	Form				**Meaning**				**Form**				**Meaning**			
	Pre-test T1		**Post-test T1**		**Pre-test T1**		**Post-test T1**		**Pre-test T1**		**Post-test T1**		**Pre-test T1**		**Post-test T1**	
	Raw	%	Raw	%	Raw	%	Raw	%	Raw	%	Raw	%	Raw	%	Raw	%
Mean	10.08	25.21	19.42	48.67	3.63	9.08	11.32	28.32	7.04	17.98	17.26	43.85	1.70	4.33	8.83	22.36
SD	5.44	13.57	7.91	19.69	3.52	8.78	6.15	15.33	5.17	13.59	6.25	16.40	2.38	6.07	4.99	12.59
Minimum	0	0	0	0	0	0	1	2.50	1	2.50	6	15	0	0	3	7.50
Maximum	22	55	33	82.50	15	37.50	27	67.50	21	52.50	28	70	10	25	18	45
Sig. differences	✓ $t(37) = -10.233$, $p = .000$				✓ $Z = -5.309$, $p = .000$				✓ $Z = -4.204$, $p = .000$				✓ $Z = -4.204$, $p = .000$			

Focusing on the relative gains for form and meaning, percentage scores are reported in Table 7.2:

Table 7.2
Relative gains' results for university students.

	Relative gains (in percentage)			
	Form		**Meaning**	
	EG (*N*=38)	CG (*N*=23)	EG (*N*=38)	CG (*N*=23)
Mean	35.22	33.14	22.88	19.52
SD	17.83	12.69	12.21	10.63
Min.	0	12.82	2.50	7.50
Max.	68.18	58.62	55.17	43.59
Sig. differences	✗		✗	

As can be observed, regardless of the condition, learners gained an average of 34% of the TWs' forms, whereas gains in meaning were lower, circa 21%. The high variability between scores is also another point that needs attention; while there were students that learned almost half of the TWs, others did not experience any improvement from the beginning to the end of the intervention. Although the means of the EG are a bit higher at the descriptive level (see Figure 7.2), inferential *t*-tests and Mann-Whitney *U* tests with relative gains as the dependent variable and experimental condition as factor revealed that differences were not large enough to reach statistical significance in word form ($t(59) = .489$, $p = .627$) and word meaning learning ($U = 358$, $z = -1.176$, $p = .239$).

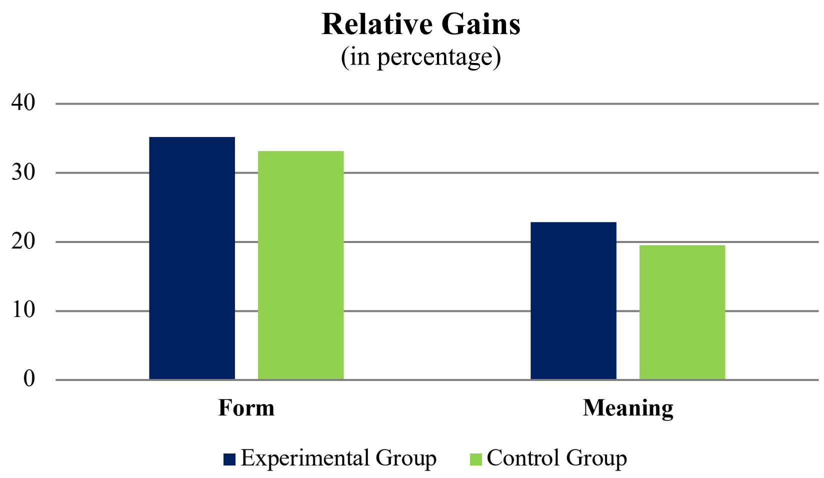


Figure 7.2 - University participants' relative gains for form and meaning, divided by condition.

7.2.2. Vocabulary learning at Grade 10.

7.2.2.1. Academic year.

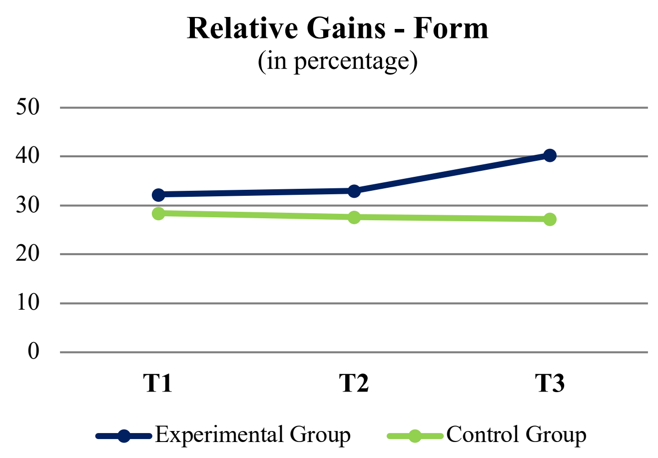
As we have previously mentioned, we will present first the analysis taking into account the results of the whole year, aimed at determining the effect of time and condition throughout all the pedagogical intervention. In order to conduct this analysis, only those learners for whom relative gains could be computed in all the three terms were selected, which resulted in a sample consisting of 24 learners in the EG (72.72% of the initial sample) and 14 in the CG (46.66%). These lower figures can be easily explained since participants for whom gains could not be calculated in one of the terms were excluded from the sample.

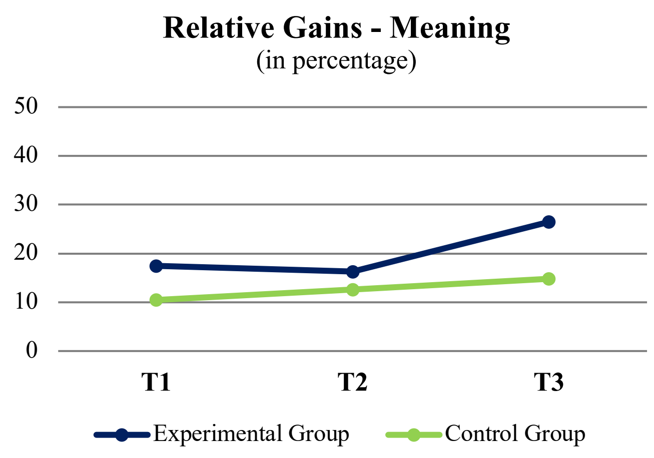
Table 7.3 provides the means, standard deviations, minimum and maximum scores of those participants for whom relative gains could be computed in the three terms. No pre- and post-tests scores are reported at this point, as the main interest of the analysis is on relative gains.

Table 7.3
Tenth graders' relative gains for form and meaning at T1, T2 and T3.

Aspect		Relative gains (in percentage)					
		Term 1		Term 2		Term 3	
		EG (*N*=24)	CG (*N*=14)	EG (*N*=24)	CG (*N*=14)	EG (*N*=24)	CG (*N*=14)
Form	Mean	32.28	28.44	32.97	27.68	40.26	27.23
	SD	18.24	14.54	17.36	11.62	20.17	11.82
	Min.	0	2.63	5.88	11.11	3.13	2.86
	Max.	62.96	53.57	76.92	52.17	80	50
Meaning	Mean	17.47	10.47	16.30	12.59	26.46	14.81
	SD	13.38	8.52	12.91	9.37	16.66	9.55
	Min.	2.50	0	2.86	0	0	0
	Max.	47.37	25	44.83	26.47	54.55	31.25

In terms of the evolution of the EG's scores across terms, there seems to be a minimal decrease between T1 and T2, whereas scores substantially increase in Term 3 (especially for the learning of word meaning). Regarding the CG, their gains in form seem to slightly decrease throughout the year, whereas the opposite tendency is observed in gains in meaning of the target items. In Figures 7.3 and 7.4, the evolution of the relative gains for both groups is visually represented.





Figures 7.3 and 7.4 - Grade 10 participants' relative gains at T1, T2 and T3 for form and meaning, divided by condition.

In order to analyse the effect of time and condition on vocabulary gains, a mixed between-within ANOVA with time as the within-subjects factor and condition as the between-subjects factor was run. So as to grasp full understanding of the phenomenon, we analyse first relative gains for form and then relative gains for meaning.

7.2.2.1.1. Word form.

Results showed that there was a non-significant main effect for time ($F(2, 72) = 1.949$, $p = .150$, partial eta squared = .051), which reveals that the relative gains of the TWs' forms did not significantly vary across the pedagogical intervention. This was later confirmed by Bonferroni corrected post-hoc tests which uncovered non-significant differences between Terms 1 and 2 ($p = 1$), Terms 1 and 3 ($p = .342$) and Terms 2 and 3 ($p = .187$). Besides, there was a non-significant main effect for condition ($F(1, 36) = 2.067$, $p = .159$, partial eta squared = .054), indicating that the number of TWs' forms learned by participants did not depend on the experimental condition they were allocated to. Finally, results showed that there was not an interaction between time and condition either ($F(2, 72) = 3.090$, $p = .052$, partial eta squared = .079), which implies that the effect of group and time altogether was not greater than the effect independently played by each factor.

7.2.2.1.2. Word meaning.

Regarding TW meanings, it was not possible to repeat the same analysis with the original relative gains as, when screened, it was seen that the variable violated some of the assumptions (e.g., normality or homogeneity of variance). In order to be able to run the mixed ANOVA, data had to be transformed and squared. After these transformations, assumptions were not violated, and we could proceed with the analysis.

A highly significant main effect for time was revealed ($F(2, 72) = 12.624$, $p = .000$, partial eta squared = .260), indicating that the number of TWs' meanings learned by the participants in both groups significantly differed throughout the intervention and that the effect size for time was considerably large (Cohen, 1988). Bonferroni post-hoc corrected coefficients showed that this difference was mostly significant between Term 3 and the other two terms (ps = .001), with the number of TWs' meanings increasing towards the end of the academic year.

Results also revealed that experimental condition was not a significant factor for the learning of TWs' meanings ($F(1, 36) = 2.836$, $p = .101$, partial eta squared = .073), and that there was not a significant interaction between time and condition ($F(2, 72) = 2.434$, $p = .095$, partial eta squared = .063). This proved that time was the only and most important factor in the analysis. In the following sections, we focus in turn on the results for each term, with larger samples of participants, starting with Term 1.

7.2.2.2. Term 1.

The final sample included a total of 57 learners, 30 of whom were randomly allocated to the EG and the remaining 27 to the CG. The descriptive statistics of both raw scores and percentages for the pre- and the post-test are found in Table 7.4. The scores on the post-test of both groups were much higher than those obtained on the pre-test, which indicates some degree of learning during Term 1. Paired samples *t*-tests and Wilcoxon signed-rank tests showed that the scores on the post-test were significantly higher than those on the pre-test for both experimental conditions and the two lexical aspects that were analysed (form and meaning).

Table 7.4
Pre- and post-test's results for Grade 10 students at T1.

	EG (N=30)								**CG** (N=27)							
	Form				**Meaning**				**Form**				**Meaning**			
	Pre-test T1		**Post-test T1**		**Pre-test T1**		**Post-test T1**		**Pre-test T1**		**Post-test T1**		**Pre-test T1**		**Post-test T1**	
	Raw	%	Raw	%	Raw	%	Raw	%	Raw	%	Raw	%	Raw	%	Raw	%
Mean	6.97	17.60	15.63	39.56	1.27	3.19	7.30	18.39	6.30	16.19	12.85	32.98	0.93	2.43	4.11	10.53
SD	3.94	9.75	7.85	19.42	2.08	5.21	6.23	15.51	4.11	10.76	7.14	18.35	1.80	4.93	3.55	8.96
Minimum	1	2.50	2	5	0	0	0	0	1	2.86	2	5.71	0	0	0	0
Maximum	18	45	30	75	9	22.50	21	52.50	16	45.71	27	68.57	8	22.86	13	32.50
Sig. differences	✓ $t(29) = -8.868, p = .000$				✓ $Z = -4.708, p = .000$				✓ $Z = -4.377, p = .000$				✓ $Z = -4.203, p = .000$			

Relative gains were computed for both form and meaning. A summary of the results can be found in Table 7.5:

Table 7.5
Relative gains' results for Grade 10 students at T1.

	Relative gains (in percentage)			
	Form		**Meaning**	
	EG (*N*=30)	CG (*N*=27)	EG (*N*=30)	CG (*N*=27)
Mean	29.75	22.75	16.39	8.94
SD	17.45	15.03	14	7.05
Min.	0	2.63	0	0
Max.	62.96	53.57	47.37	25
Sig. differences	✗		✓	

As shown by the descriptive results, the EG gained more TWs' forms than the CG, the difference being of seven percentage points. However, this did not reach statistical significance ($t(55) = 1.613$, $p = .112$). In the case of TWs' meanings, the difference between the EG and CG was of 7.45 percentage points, and a Mann-Whitney *U* test yielded significant differences between both experimental conditions ($U = 278.5$, $z = -2.029$, $p = .042$).

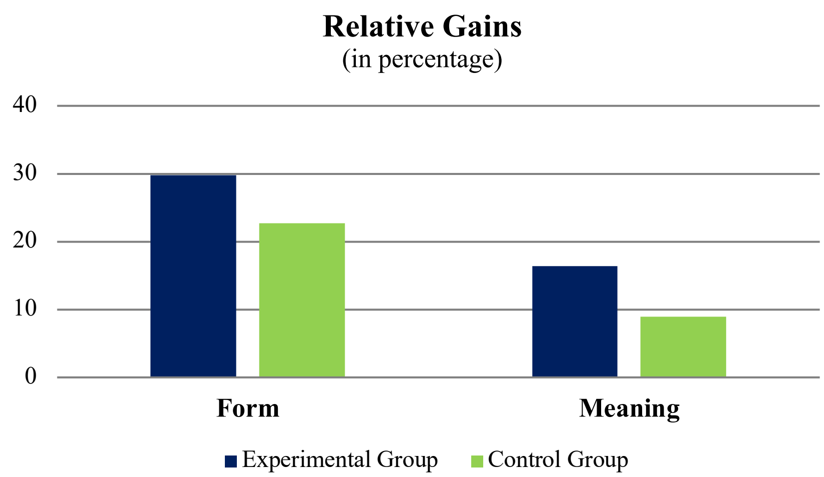


Figure 7.5 - Grade 10 participants' relative gains at T1 for form and meaning, divided by condition.

7.2.2.3. Term 2.

In Term 2, the sample included in the final analysis was composed of 52 participants, 29 who were additionally exposed to multimodal input and 23 who followed traditional curricular instruction only. In terms of the descriptive statistics, scores are presented in raw numbers and in percentages in Table 7.6. As was the case in Term 1, the TWs' forms showed the largest gains from the beginning to the end of the academic term in both groups, and the differences were statistically significant as shown by paired samples *t*-tests. The same applied to word meanings, as Wilcoxon signed-rank tests showed that statistically significant progress was made from the start to the end of the term, regardless of the experimental condition (see Table 7.6).

Table 7.6
Pre- and post-test's results for Grade 10 students at T2.

	EG (*N*=29)								**CG** (*N*=23)							
	Form				**Meaning**				**Form**				**Meaning**			
	Pre-test T2		**Post-test T2**		**Pre-test T2**		**Post-test T2**		**Pre-test T2**		**Post-test T2**		**Pre-test T2**		**Post-test T2**	
	Raw	%	Raw	%	Raw	%	Raw	%	Raw	%	Raw	%	Raw	%	Raw	%
Mean	8.45	24.63	16.28	47.69	0.90	2.58	5.66	16.49	7.43	22.11	12.09	36.07	0.39	1.20	3.48	10.50
SD	4.83	13.67	7.15	20.38	1.97	5.63	4.99	14.61	4.33	12.50	6.88	19.42	0.89	2.76	3.27	9.83
Minimum	1	2.86	3	8.57	0	0	1	2.86	1	3.33	1	3.33	0	0	0	0
Maximum	21	60	29	82.86	8	22.86	18	51.43	14	40	25	71.43	3	10	10	28.57
Sig. differences	✓ $t(28) = -9.946, p = .000$				✓ $Z = -4.710, p = .000$				✓ $t(22) = -5.724, p = .000$				✓ $Z = -3.731, p = .000$			

To analyse the effect of viewing on the learning of TWs' forms and meanings, relative gains were computed for both groups (see Table 7.7). The descriptive statistics of the tenth graders' relative gains in Term 2 showed that the largest differences between groups were found in the formal aspect of the TWs (+10.53 points in favour of the EG). This was corroborated by an independent samples *t*-test, which showed that sustained exposure to captioned TV viewing was beneficial for word form acquisition, with the EG significantly outperforming the CG ($t(50) = 2.465$, $p = .017$). However, the difference in the number of TWs' meanings learned by the two groups did not reach statistical significance ($U = 259$, $z = -1.377$, $p = .169$), even if descriptive results showed that the EG's relative gains were 5.33% higher.

Table 7.7
Relative gains' results for Grade 10 students at T2.

	Relative gains (in percentage)			
	Form		**Meaning**	
	EG (*N*=29)	CG (*N*=23)	EG (*N*=29)	CG (*N*=23)
Mean	33.55	23.02	15.13	9.80
SD	16.53	13.55	12.34	8.68
Min.	5.88	3.03	2.86	0
Max.	76.92	52.17	44.83	26.47
Sig. differences	✓		✗	

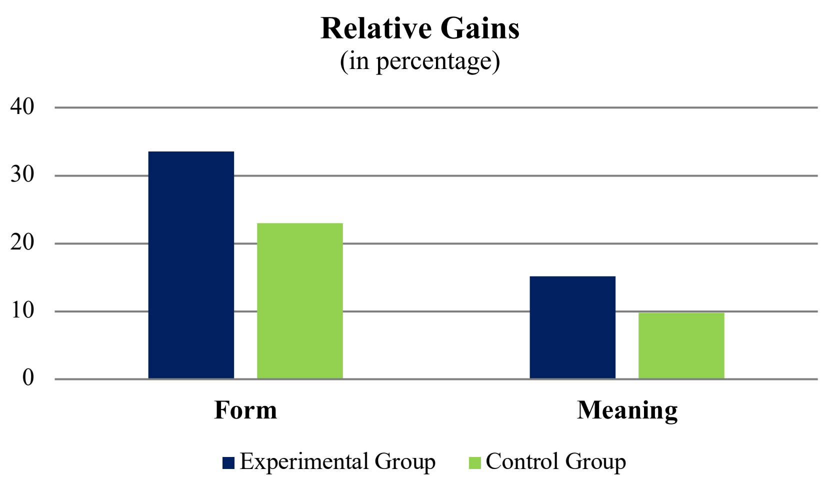


Figure 7.6 - Grade 10 participants' relative gains at T2 for form and meaning, divided by condition.

7.2.2.4. Term 3.

In Term 3, the final sample comprised 48 students, 28 from the EG and 20 from the CG. Table 7.8 provides the means, standard deviations, and minimum and maximum scores. As in the previous terms, both the EG and CG learned new vocabulary in Term 3, as shown by paired samples *t*-tests for word form and Wilcoxon signed-rank tests for word meaning: the scores on the post-tests were significantly higher than the scores on the pre-test for both groups and the two lexical aspects (see Table 7.8), and the effect sizes indicated that progress was considerable.

Table 7.8
Pre- and post-test's results for Grade 10 students at T3.

	EG (*N*=28)								**CG** (*N*=20)							
	Form				**Meaning**				**Form**				**Meaning**			
	Pre-test T3		**Post-test T3**		**Pre-test T3**		**Post-test T3**		**Pre-test T3**		**Post-test T3**		**Pre-test T3**		**Post-test T3**	
	Raw	%	Raw	%	Raw	%	Raw	%	Raw	%	Raw	%	Raw	%	Raw	%
Mean	8.93	26.80	17.07	50.39	2.71	8.28	10.07	24.92	10.35	30.88	14.30	42.57	2	5.83	6.10	14.55
SD	5.17	16.03	7.88	22.59	2.69	8.76	6.31	18.54	5.99	17.08	7.64	21.53	2.20	6.38	4.47	13.88
Minimum	1	2.86	4	11.43	0	0	0	0	0	0	1	2.86	0	0	0	0
Maximum	20	66.67	31	88.57	12	40	21	60	20	57.14	25	73.33	8	22.86	15	42.86
Sig. differences	✓ $t(27) = -10.139$, $p = .000$				✓ $Z = -3.924$, $p = .000$				✓ $t(27) = -5.152$, $p = .000$				✓ $Z = -3.086$, $p = .002$			

In terms of relative gains, Table 7.9 shows the descriptive scores for the two groups and Figure 7.7 the graphic representation. The EG learned 11.26% more of TWs' forms than the CG, being this difference large enough to reach statistical significance, as shown by an independent samples *t*-test ($t(46) = 2.198$, $p = .033$). The difference between word meaning gains was of 10.58 percentage points (in favour of the EG), and inferential tests also showed that it was statistically significant ($t(44.773) = 2.821$, $p = .007$). In light of these results, differences in Term 3 are larger than in any of the other two terms already presented in the previous sections.

Table 7.9
Relative gains' results for Grade 10 students at T3.

	Relative gains (in percentage)			
	Form		**Meaning**	
	EG (*N*=28)	CG (*N*=20)	EG (*N*=28)	CG (*N*=20)
Mean	38.64	27.38	25.11	14.53
SD	19.43	14.33	16.22	9.66
Min.	3.13	2.86	0	0
Max.	80	52.94	54.55	31.25
Sig. differences	✓		✓	

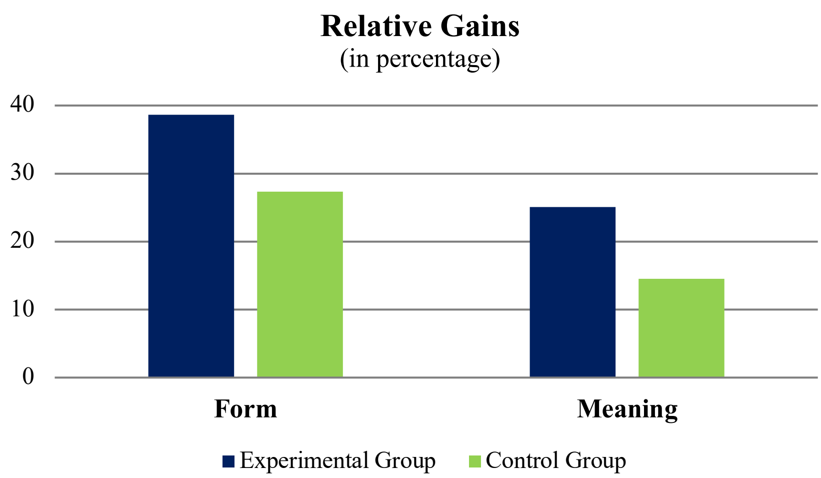


Figure 7.7 - Grade 10 participants' relative gains at T3 for form and meaning, divided by condition.

7.2.3. Vocabulary learning at Grade 6.

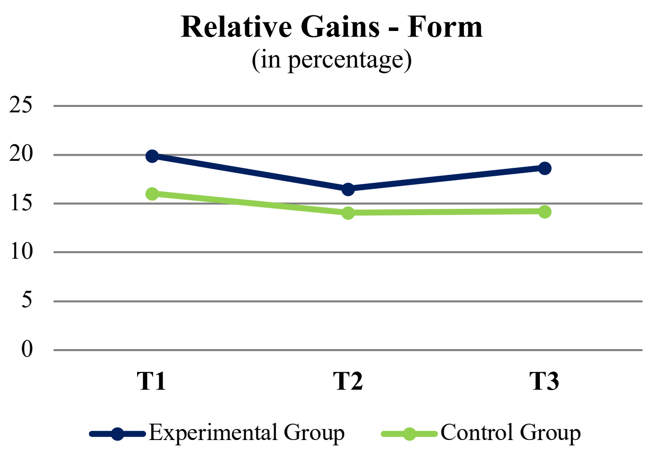
In this section, data collected in Grade 6 will be analysed following the same procedure in Grade 10: first, by analysing the overall effect of time and condition on the results and then by focusing on each term in turn, considering always the learning of word forms and word meanings.

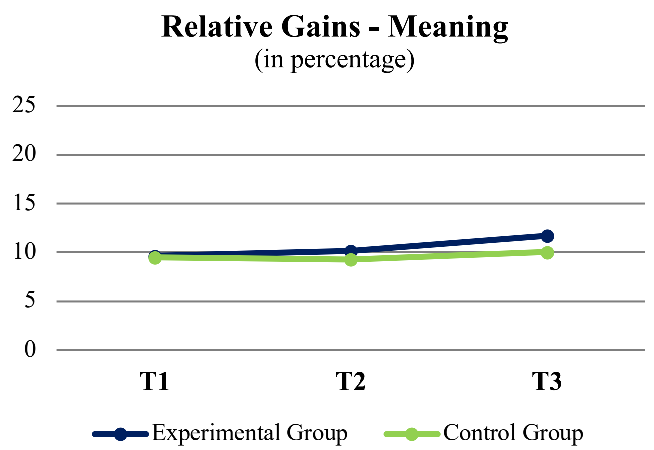
7.2.3.1. Academic year.

To conduct this analysis, the performance of 29 participants was studied longitudinally all along the academic year. These included 18 learners from the EG and 11 from the CG (66.66% of the initial sample in the case of the EG and a 44% in the case of the CG). As can be seen in Table 7.10 and Figures 7.8 and 7.9, where the descriptive statistics for the relative gains across the three terms are reported and graphically represented, the results of this subsample follow the general pattern already seen in Grade 10. In terms of their evolution, the relative gains for TWs' forms decrease between T1 and T2 and increase again at the end of the pedagogical intervention. Regarding the other measure (word meaning), there is a steady increase from the beginning to the end of the academic year.

Table 7.10
Sixth graders' relative gains for form and meaning at T1, T2 and T3.

Aspect		**Relative gains** (in percentage)					
		Term 1		**Term 2**		**Term 3**	
		EG (*N*=18)	CG (*N*=11)	EG (*N*=18)	CG (*N*=11)	EG (*N*=18)	CG (*N*=11)
Form	Mean	19.91	16.04	16.53	14.07	18.69	14.24
	SD	11.69	7.07	10.39	7.33	11.76	9.82
	Min.	2.86	5	2.70	5.26	0	2.56
	Max.	43.33	32.35	41.94	28.57	48.48	39.39
Meaning	Mean	9.66	9.53	10.17	9.28	11.71	10.06
	SD	8.67	9.46	6.20	10.55	7.38	6.83
	Min.	0	0	5	0	2.50	5.13
	Max.	26.32	28.21	23.08	28.95	33.33	28.95





Figures 7.8 and 7.9 - Grade 6 participants' relative gains at T1, T2 and T3 for form and meaning, divided by condition.

In the sections that follow, the results of two mixed between-within ANOVAs (one for each measure -form and meaning-) with time as the within-subject factor and experimental condition as the between-groups factor are presented.

7.2.3.1.1. Word form.

It should be first noted that data presented a normal distribution in all but one of the academic terms (Term 3) following the Shapiro-Wilk test. However, as mixed ANOVAs are quite tolerant to violations of this assumption (Pallant, 2013), it was decided to run this statistical test. Besides, no other assumptions were violated.

Results showed a non-significant main effect for time ($F(2, 54) = 1.517$, $p = .229$, partial eta squared = .053), and pairwise comparisons also revealed no significant differences between any of the terms. The between-group comparison also mirrored the within-group analysis, as no significant main effect for condition was found ($F(1, 27) = 1.056$, $p = .313$, partial eta squared = .038). Finally, and as expected after seeing the strength of the two main effects, the interaction between time and condition was not significant either ($F(2, 54) = .223$, $p = .801$, partial eta squared = .008).

7.2.3.1.2. Word meaning.

In the preliminary analysis of the data, it was clearly shown that this variable was not normally distributed and skewed to the left. Data had to be transformed and, from the different options available, the logarithm function proved to be the best solution to the non-normally distributed data. After the transformation was applied, data was normally distributed in five of the six groups in which it was divided, and it almost reached a normal distribution in the sixth. As stated before, the mixed ANOVA is quite robust to the violation of this assumption, so the fact that one group did not meet it would not affect the overall result of the test (Pallant, 2013). The rest of the assumptions were not violated when data was transformed.

Time was found to be a significant main effect for the learning of word meanings ($F(2, 54) = 3.720$, $p = .031$, partial eta squared = .121). However, none of the Bonferroni pairwise comparisons between any of the terms gave a significant result (T1 vs. T2 → $p = 1$ / T2 vs. T3 → $p = .076$ / T1 vs. T3 → $p = .086$). There was not a significant main effect for condition on the learning of the TWs' meanings, as the scores from the two groups were very similar in two of the terms ($F(1, 27) = .619$, $p = .438$, partial eta squared = .022). Finally, the time*group interaction was also non-significant ($F(2, 54) = 1.487$, $p = .235$, partial eta squared = .052).

Next, in sections 7.2.3.2., 7.2.3.3. and 7.2.3.4., data from Grade 6 learners will be analysed term-independently with larger numbers of participants in each of the terms.

7.2.3.2. Term 1.

In Term 1, 40 learners fulfilled the criteria to be considered valid participants for our analysis; 22 of them belonged to the EG and 18 to the CG. As in the two previous proficiency groups, first, the descriptive statistics, in both raw scores and percentages, will be reported, followed by the inferential tests that help us to answer the first RQ. As can be seen in Table 7.11, scores' variability was lower in Grade 6 than in Grade 10. Nonetheless, this was probably caused by the fact that the scores were also lower and the participants' level more homogeneous.

The scores on the pre-test were always lower than on the post-test, which indicates that progress was made during Term 1. Wilcoxon signed-rank tests confirmed that the progress observed in the descriptive results was statistically significant in all conditions and for both acquisition of word forms and word meanings (please check Table 7.11), and that this was remarkable, as shown by the large effect sizes (*r*s ranging from .61 to .85).

Table 7.11
Pre- and post-test's results for Grade 6 students at T1.

	EG (*N*=22)								**CG** (*N*=18)							
	Form				**Meaning**				**Form**				**Meaning**			
	Pre-test T1		**Post-test T1**		**Pre-test T1**		**Post-test T1**		**Pre-test T1**		**Post-test T1**		**Pre-test T1**		**Post-test T1**	
	Raw	%	Raw	%	Raw	%	Raw	%	Raw	%	Raw	%	Raw	%	Raw	%
Mean	4.14	10.44	10.50	26.49	3.22	8.17	9.06	22.72	1.45	3.65	5.23	13.10	1.83	4.62	5.39	14.14
SD	3.03	7.50	5.41	13.38	3.23	8.08	5.91	14.86	1.90	4.74	3.93	9.79	2.50	6.26	5.39	13.48
Minimum	1	2.50	1	2.86	0	0	2	5	0	0	0	0	0	0	0	0
Maximum	12	30	23	57.50	12	30	22	55	7	17.50	14	35	9	22.50	18	45
Sig. differences	✓ $Z = -4.013$, $p = .000$				✓ $Z = -3.829$, $p = .000$				✓ $Z = -3.728$, $p = .000$				✓ $Z = -3.421$, $p = .001$			

Regarding the relative gains (see Table 7.12 for the results summary and Figure 7.10 for a visual representation), the EG gained 2.43 percentage points more of word forms than the CG, although this difference was not statistically significant, as shown by an independent samples *t*-test ($t(38) = .721$, $p = .475$). The CG gained more TWs' meanings (+0.47 points) than the EG, which had not happened before in any of the terms in the two proficiency groups already analysed. However, a Mann-Whitney *U* test showed that this did not actually reach statistical significance ($U = 194$, $z = -.109$, $p = .913$).

Table 7.12
Relative gains' results for Grade 6 students at T1.

	Relative gains (in percentage)			
	Form		**Meaning**	
	EG (*N*=22)	CG (*N*=18)	EG (*N*=22)	CG (*N*=18)
Mean	20.82	18.39	10.33	10.80
SD	10.75	10.42	8.35	10.62
Min.	2.86	5	0	0
Max.	43.33	41.18	26.32	31.25
Sig. differences	✘		✘	

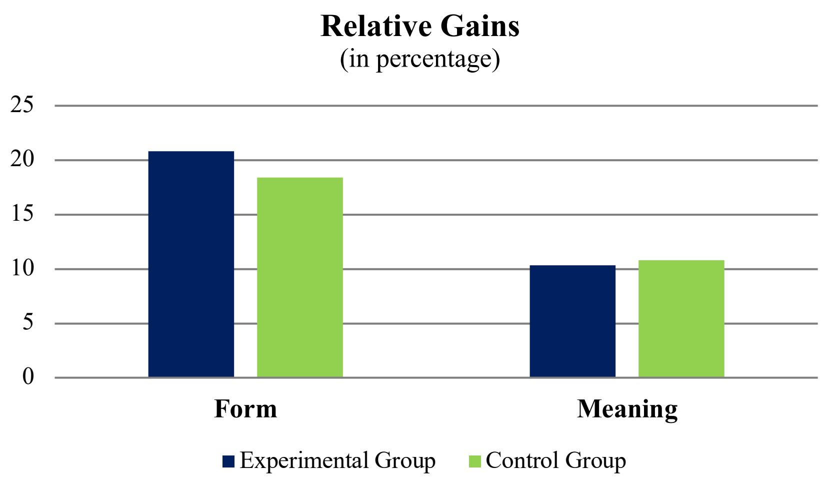


Figure 7.10 - Grade 6 participants' relative gains at T1 for form and meaning, divided by condition.

7.2.3.3. Term 2.

In Term 2, the analysis was conducted with 40 students: 23 of whom belonging to the experimental condition and the remaining 17 to the CG. Table 7.13 shows the results of the pre- and post-test in Term 2. As expected, higher scores were obtained at the end of the term than at the beginning in both experimental conditions. Paired samples *t*-tests revealed that the scores for word form learning were significantly higher on the post-test than on the pre-test in the two experimental conditions. The same applies to word meaning learning, as shown by non-parametric Wilcoxon signed-rank tests. In all cases, the effect size was large (*r*s varying from .82 to 1.53 in the EG and between .71 and 1.55 in the CG).

Table 7.13
Pre- and post-test's results for Grade 6 students at T2.

	EG (*N*=23)								**CG** (*N*=17)							
	Form				**Meaning**				**Form**				**Meaning**			
	Pre-test T2		**Post-test T2**		**Pre-test T2**		**Post-test T2**		**Pre-test T2**		**Post-test T2**		**Pre-test T2**		**Post-test T2**	
	Raw	%	Raw	%	Raw	%	Raw	%	Raw	%	Raw	%	Raw	%	Raw	%
Mean	4.96	12.59	10.13	25.65	1.35	3.42	5.61	14.19	3.88	10.21	7.53	19.75	1.88	4.92	5	13.05
SD	1.80	4.45	4.59	11.29	1.23	3.06	2.94	7.25	1.83	5.04	3.59	9.37	1.58	4.23	4.33	11.16
Minimum	2	5	4	10	0	0	2	5	2	5	2	5	0	0	0	0
Maximum	9	22.50	22	55	5	12.50	13	32.50	8	22.86	15	37.50	5	14.29	15	37.50
Sig. differences	$t(22) = -7.318, p = .000$				$Z = -4.210, p = .000$				$t(16) = -6.382, p = .000$				$Z = -3.551, p = .000$			

Focusing on the gains during the second term, we can see that the scores of the EG were higher than those of the CG. Regarding word forms, the difference between the two conditions was of 5.01 percentage points. Nevertheless, it did not reach statistical significance ($t(38) = 1.687$, $p = .100$). The number of word meanings learned was also higher in the EG than in the CG (+1.82 points), but the difference was not statistically significant either, as shown by a Mann-Whitney *U* test ($U = 135.5$, $z = -1.643$, $p = .100$) (see Table 7.14 and Figure 7.11).

Table 7.14
Relative gains' results for Grade 6 students at T2.

	Relative gains (in percentage)			
	Form		**Meaning**	
	EG (*N*=23)	CG (*N*=17)	EG (*N*=23)	CG (*N*=17)
Mean	18.08	13.07	11.30	9.48
SD	10.49	7.30	6.22	9.35
Min.	2.70	0	5	0
Max.	41.94	28.57	23.08	28.95
Sig. differences	✘		✘	

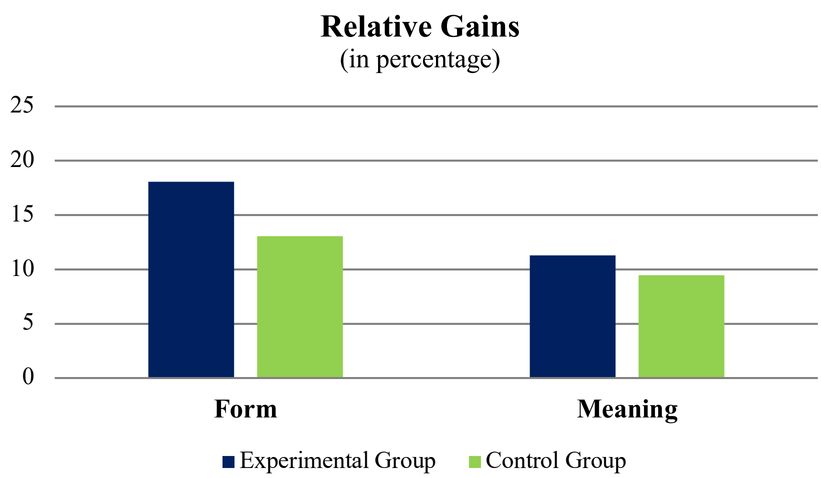


Figure 7.11 - Grade 6 participants' relative gains at T2 for form and meaning, divided by condition.

7.2.3.4. Term 3.

In the last term, 40 participants were included in the final analysis: 22 belonged to the EG and 18 to the CG. Regarding the descriptive results (see Table 7.15), a set of Wilcoxon signed-rank tests and a paired samples *t*-test revealed that the scores on word forms and word meanings by both the EG and the CG were significantly higher on the post-test as compared to the pre-test (ps $\leq .001$ in all cases). Progress was considerable, as suggested by the effect sizes (rs ranging from .84 to 1.64 in the EG and from .66 to .85 in the CG).

Table 7.15
Pre- and post-test's results for Grade 6 students at T3.

	EG (N=22)								CG (N=18)							
	Form				Meaning				Form				Meaning			
	Pre-test T3		Post-test T3		Pre-test T3		Post-test T3		Pre-test T3		Post-test T3		Pre-test T3		Post-test T3	
	Raw	%	Raw	%	Raw	%	Raw	%	Raw	%	Raw	%	Raw	%	Raw	%
Mean	6.05	15.11	11.86	29.66	2.91	7.27	7.09	17.73	3.22	8.59	6.50	17.06	1.61	4.29	4.44	11.63
SD	3.18	7.96	5.73	14.32	1.48	3.69	3.34	8.34	2.42	6.68	4.22	10.69	1.61	4.54	3.01	7.86
Minimum	1	2.50	3	7.50	0	0	1	2.50	0	0	2	5	0	0	0	0
Maximum	13	32.50	23	57.50	5	12.50	16	40	10	28.57	20	50	7	20	13	32.50
Sig. differences	✓ $t(21) = -7.678, p = .000$				✓ $Z = -4.134, p = .000$				✓ $Z = -3.412, p = .001$				✓ $Z = -3.623, p = .000$			

Regarding the relative gains, the difference in the number of TWs' forms learned by both experimental conditions was of 7.91 percentage points. This difference resulted to be statistically significant following a Mann-Whitney *U* test, indicating that the relative gains for form obtained by the EG in Term 3 were significantly higher than those of the CG (U = 103.5, z = -2.570, p = .010). In relation to word meaning, the difference between conditions was of 3.23 points, favouring the EG, and another Mann-Whitney *U* test further showed that such difference was statistically significant (U = 118.5, z = -2.164, p = .030) (see Table 7.16 and Figure 7.12).

Table 7.16
Relative gains' results for Grade 6 students at T3.

	Relative gains (in percentage)			
	Form		**Meaning**	
	EG (*N*=22)	CG (*N*=18)	EG (*N*=22)	CG (*N*=18)
Mean	20.19	12.28	11.70	8.47
SD	11.43	8.59	7.05	5.95
Min.	0	2.56	2.50	0
Max.	48.48	39.39	33.33	28.95
Sig. differences	✓		✓	

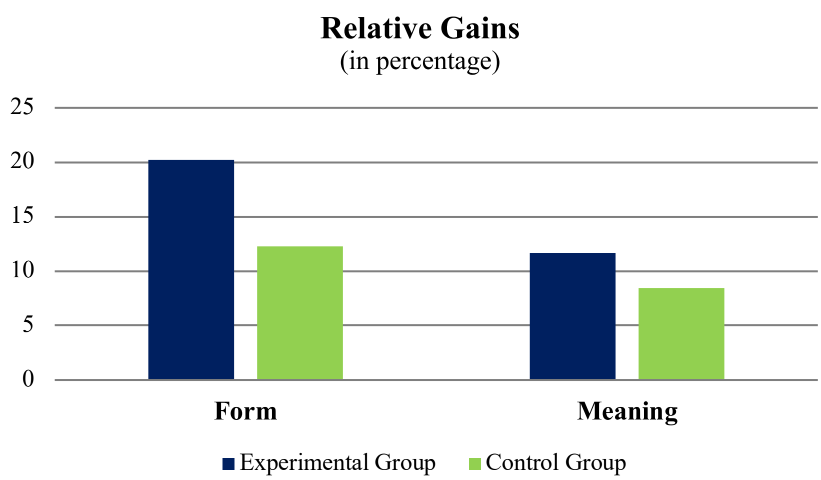


Figure 7.12 - Grade 6 participants' relative gains at T3 for form and meaning, divided by condition.

7.2.4. Summary of the results for RQ1.

Regarding the results from the first RQ, tapping into the effects of video viewing on vocabulary learning, we first see that, at university, participants from both experimental conditions (EG and CG) made significant progress from the start to the end of the intervention, lasting one academic term. However, although descriptive statistics pointed towards derisory differences between conditions, these were not later confirmed by inferential tests. The results of the independent samples *t*-tests and Mann-Whitney *U* tests showed that condition was not a determinant factor in the learning of new vocabulary in freshman university EFL learners.

Table 7.17
Summary of the findings for RQ1 at university.

Aspect	**Term - T1**
Form	✘ (p = .627)
Meaning	✘ (p = .239)

Talking about the overall effect of time and condition on the whole intervention in Grade 10 and Grade 6 (see Table 7.18), time played a significant role in the learning of TWs' meanings (not in the learning of word forms). This showed that, regardless of the experimental condition, learners' scores for word meaning learning varied from the beginning to the end of the academic year, being those at T3 usually higher than at T1. The experimental condition participants were allocated to did not prove to be a significant factor, neither for word form nor for word meaning learning in any of the two proficiency groups. Finally, the time*condition interaction also failed to reach statistical significance.

Table 7.18
Summary of the results of the mixed ANOVAs in Grade 10 and Grade 6.

	Aspect	**Factor**		
		Time	**Condition**	**Time * Condition Interaction**
Grade 10	Form	✗ (*p* = .150)	✗ (*p* = .159)	✗ (*p* = .052)
	Meaning	✓ (*p* = .000)	✗ (*p* = .101)	✗ (*p* = .095)
Grade 6	Form	✗ (*p* = .229)	✗ (*p* = .313)	✗ (*p* = .801)
	Meaning	✓ (*p* = .031)	✗ (*p* = .438)	✗ (*p* = .235)

When data was analysed term-independently in Grade 10 and Grade 6, it should be noted that statistically significant progress from the beginning to the end of the term was a common finding and for both EG and CG. However, significant differences between experimental conditions depended on the term, grade and lexical aspect being analysed. In Term 1, no differences were found in Grade 6, although gains in TWs' meanings (not in word forms) did reach statistical significance in Grade 10. A similar pattern was observed in Term 2, when no significant differences were found in Grade 6, but they were found in Grade 10 for the learning of TWs' forms (not in word meaning learning, though). Finally, in Term 3, differences were found in the two lexical aspects studied (form and meaning) and in both Grade 10 and Grade 6. It is important to bear in mind that, in those cases where differences were significant, they were always in favour of the experimental condition, who had been additionally exposed to multimodal input (see Table 7.19 for a summary of the results).

Table 7.19
Summary of the findings for RQ1 in Grade 10 and Grade 6.

	Aspect	Term		
		T1	T2	T3
Grade 10	Form	✗ (p = .112)	✓ (p = .017)	✓ (p = .033)
	Meaning	✓ (p = .042)	✗ (p = .169)	✓ (p = .007)
Grade 6	Form	✗ (p = .475)	✗ (p = .100)	✓ (p = .010)
	Meaning	✗ (p = .913)	✗ (p = .100)	✓ (p = .030)

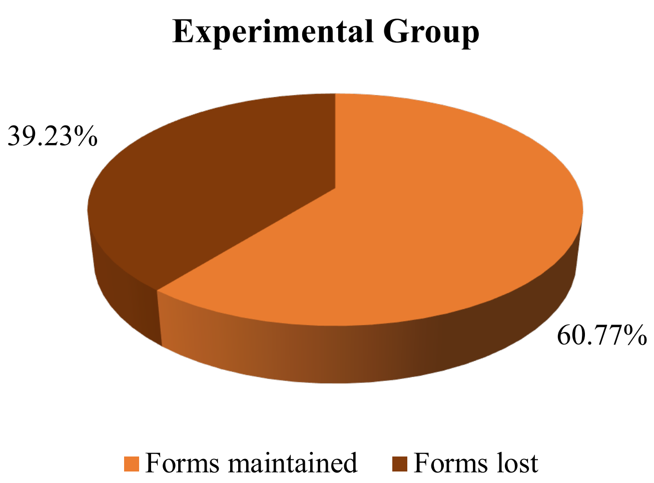
7.3. RQ1.1. - Vocabulary retention after extensive viewing

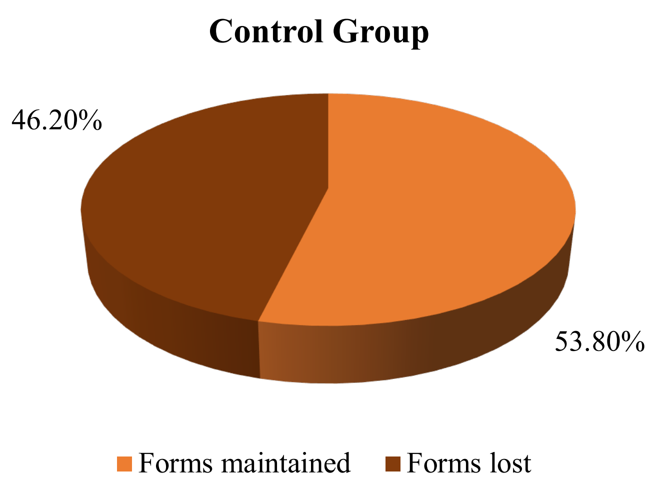
This section presents the results for RQ1.1., which analyses the long-term retention of the vocabulary which was learned during the last term of the pedagogical intervention in Grade 10 and Grade 6. Results will be presented first for high-school learners and then for primary school students.

7.3.1. Vocabulary retention at Grade 10.

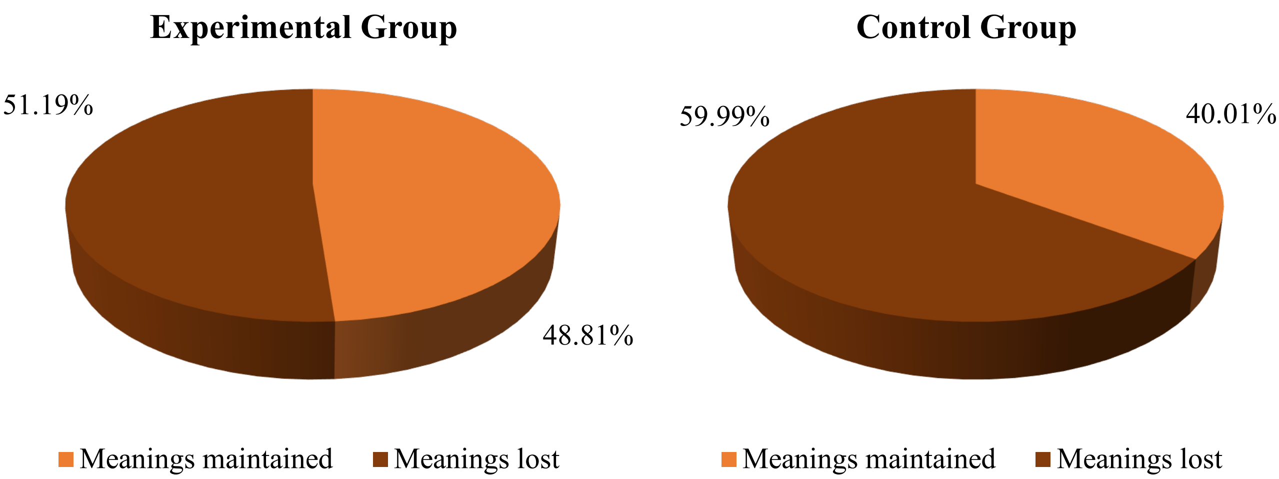
At this level, 40 students who had completed the T3 pre- and post-tests also took the delayed post-test: 24 belonged to the EG and 16 to the CG. Table 7.20 shows that almost 61% of the TWs' forms learned during the last term of the intervention were still remembered by the EG on the delayed test, whereas the CG maintained 53.80% of the word forms acquired eight months before. The percentage of word meanings still remembered by the EG was of 48.41% while the CG remembered 35.18% of the TWs' meanings previously acquired. It is also worth noting that some participants in both groups were able to remember all the word forms and meanings they had previously learned while others did not remember any.

In terms of the between-group comparison, even if the EG remembered more forms (+6.97%) and meanings (+13.07%) than the CG, an independent samples *t*-test revealed no significant differences in the percentage of the TWs' forms remembered ($t(38) = .755$, $p = .455$), indicating that modality of input (i.e., having been additionally exposed to L2 subtitled TV series) did not affect the number of word forms retained in the long term. Focusing on TWs' meanings, *t*-test results showed no significant differences in the percentage of word meanings correctly remembered on the delayed test, concluding that modality of input did not have an effect on the retention rate ($t(37) = 1.255$, $p = .217$). No inferential tests were conducted with the percentage of word forms and word meanings lost because we were mainly interested in how much learners retained. Besides, the results would not add any extra information, as the two measures (knowledge maintained and lost) tap into the same construct and the sum of the two percentages must be 100% in all cases.





Figures 7.13 and 7.14 - Percentage of word forms maintained and lost on the delayed test in Grade 10.

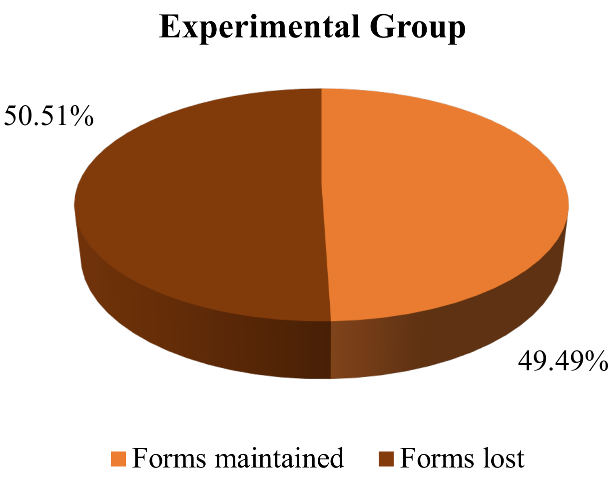


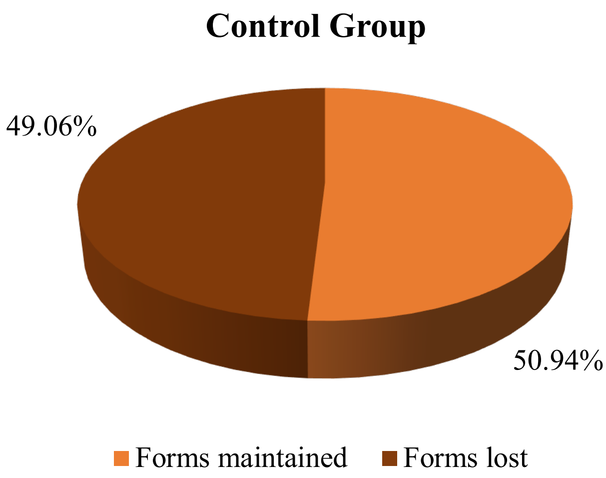
Figures 7.15 and 7.16 - Percentage of word meanings maintained and lost on the delayed test in Grade 10.

7.3.2. Vocabulary retention at Grade 6.

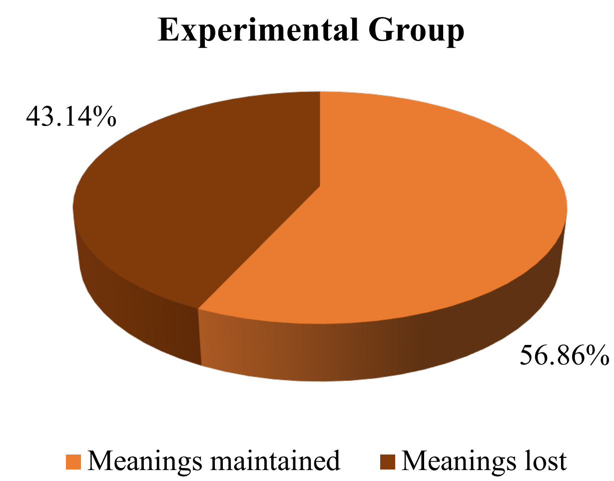
In Grade 6, data from 34 EFL learners was thoroughly analysed: 19 students belonged to the EG while the remaining 15 belonged to the CG. Table 7.21 indicates that around half of the TW forms learned during the intervention were still remembered on the delayed test by both the EG and the CG. Word meanings were better retained, since both experimental conditions still remembered a 57% of the TW meanings learned during the intervention. High variability was also seen at this proficiency level, as some students were able to remember all the word forms and word meanings they had learned eight months before while others did not remember any.

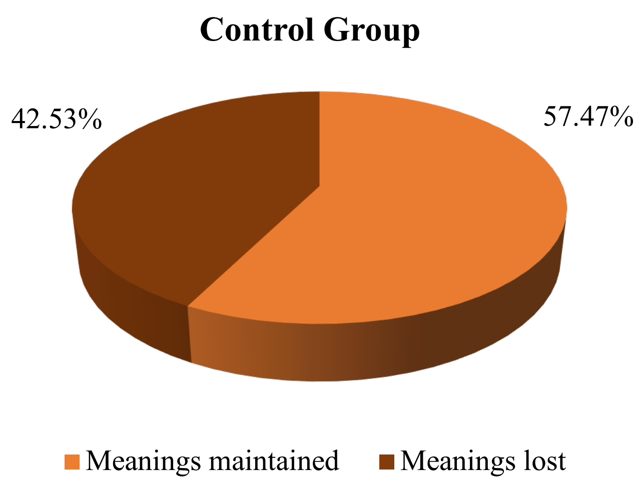
If the two groups are compared, the differences between conditions on retention of word forms were minimal, and this was confirmed by an independent-samples *t*-test, which showed that the percentage of word forms retained on the delayed test did not vary across experimental conditions ($t(31) = -.137$, $p = .892$). Regarding the TWs' meanings, a Mann-Whitney *U* test was conducted and revealed no significant differences between the percentage of word meanings remembered on the delayed test by both the EG and the CG (U = 124.5, z = -.313, p = .754). Similarly to Grade 10, no inferential tests were conducted with the percentage of TW forms and meanings lost.





Figures 7.17 and 7.18 - Percentage of word forms maintained and lost on the delayed test in Grade 6.





Figures 7.19 and 7.20 - Percentage of word meanings maintained and lost on the delayed test in Grade 6.

Table 7.20
Raw and percentage scores of word forms and word meanings maintained and lost by Grade 10 learners on the delayed test.

	Form								**Meaning**							
	Maintained				**Lost**				**Maintained**				**Lost**			
	EG (*N*=24)		CG (*N*=16)		EG (*N*=24)		CG (*N*=16)		EG (*N*=24)		CG (*N*=16)		EG (*N*=24)		CG (*N*=16)	
	Raw	%	Raw	%	Raw	%	Raw	%	Raw	%	Raw	%	Raw	%	Raw	%
Mean	5.50	60.77	3.56	53.80	3.58	39.23	2.63	46.20	4.29	48.81	1.94	35.18	3.13	51.19	2.44	64.82
SD	3.28	27.54	2.78	30.20	2.52	27.54	1.82	30.20	3.91	34.09	1.84	32.24	2.52	34.09	1.63	32.24
Minimum	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Maximum	10	100	11	100	9	100	6	100	12	100	5	100	10	100	5	100
Sig. differences	✗								✗							

Table 7.21
Raw and percentage scores of word forms and word meanings maintained and lost by Grade 6 learners on the delayed test.

	Form								**Meaning**							
	Maintained				**Lost**				**Maintained**				**Lost**			
	EG (*N*=19)		CG (*N*=15)		EG (*N*=19)		CG (*N*=15)		EG (*N*=19)		CG (*N*=15)		EG (*N*=19)		CG (*N*=15)	
	Raw	%	Raw	%	Raw	%	Raw	%	Raw	%	Raw	%	Raw	%	Raw	%
Mean	3.53	49.49	2.20	50.94	3.58	50.51	2	49.06	2.58	56.86	1.93	57.47	1.95	43.14	1.27	42.53
SD	2.39	26.30	1.82	34.55	2.09	26.30	1.60	34.55	2.19	27.76	1.83	39.02	1.31	27.76	1.44	39.02
Minimum	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Maximum	11	100	7	100	8	100	5	100	10	100	6	100	5	100	5	100
Sig. differences	✗								✗							

7.3.3. Summary of the results for RQ1.1.

RQ 1.1. enquired about vocabulary retention on the delayed test, eight months after the end of the intervention. Long-term retention could be studied in Grade 6 and Grade 10 learners, who were still at the same institutions the academic year after the experiment, but not in university students.

As Table 7.22 shows, modality of input (i.e., whether or not learners had been exposed to subtitled TV series) did not have a significant effect on the retention of the vocabulary learned during the third term of the intervention. In other words, when only the TWs learned in Term 3 were considered (excluding those already known at the beginning of the term and those unlearned during T3), the EGs did not show benefits in terms of word form or word meaning retention.

Table 7.22
Summary of the results for RQ1.1.

Measure	Level	Aspect	
		Form	**Meaning**
Percentage of TWs maintained	Grade 10	✘ (p = .455)	✘ (p = .217)
	Grade 6	✘ (p = .892)	✘ (p = .754)

7.4. RQ2 - Content comprehension of subtitled TV series

This section outlines the results of the second RQ, which looks into the content comprehension of subtitled TV series in the three proficiency groups analysed: university, Grade 10 and Grade 6 EFL learners (in this same order).

7.4.1. Content comprehension at university.

Figure 7.21 shows a visual representation of university learners' mean scores per episode and Table 7.23 presents the descriptive statistics of the comprehension scores obtained by this proficiency group (*N*=38 in most cases).

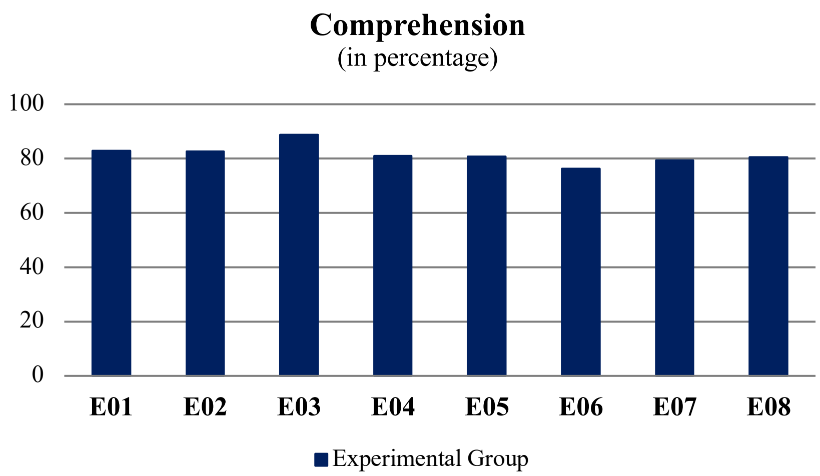


Figure 7.21 - Comprehension scores of university participants throughout the academic term.

Table 7.23
Comprehension tests' descriptive statistics for university students.

	I Love Lucy (in percentage)							
	E01	**E02**	**E03**	**E04**	**E05**	**E06**	**E07**	**E08**
N	38	38	38	38	38	38	37	38
Mean	82.89	82.57	88.65	80.92	80.76	76.32	79.22	80.43
SD	11.60	13.87	8.70	14.96	13.58	11.55	12.59	15.32
Min.	43.75	50	62.50	31.25	31.25	43.75	56.25	37.50
Max.	100	100	100	100	100	93.75	100	100

In the first place, it needs to be said that the mean comprehension score obtained by this group was 81.47%, which converts into a mean of 13.04 points on the comprehension tests. The minimum score was achieved in E06 (76.32%), whereas E03 was the easiest episode to understand, with a score of 88.65%. It also needs to be acknowledged that comprehension scores did not remain stable throughout the academic term. This was further ratified by a Friedman test ($N = 37$), which showed that they significantly varied in Term 1 at the university level, being highly episode-dependent and without a clear ascending or descending pattern ($\chi^2(7) = 27.35$, $p = .000$). Post-hoc comparisons showed that significant differences were observed between E03 and E06 ($p = .000$) and E03 vs. E07 ($p = .019$), revealing that the scores on the third episode were significantly higher than in the other two episodes.

7.4.2. Content comprehension at Grade 10.

Figure 7.22 below shows the comprehension scores of Grade 10 learners throughout the academic year. The two dotted lines between E08 and E09 and E15 and E16 mark the end of one term and the beginning of the next; besides, the first dotted line also marks the change of the TV series (from *I Love Lucy* to *Seinfeld*).

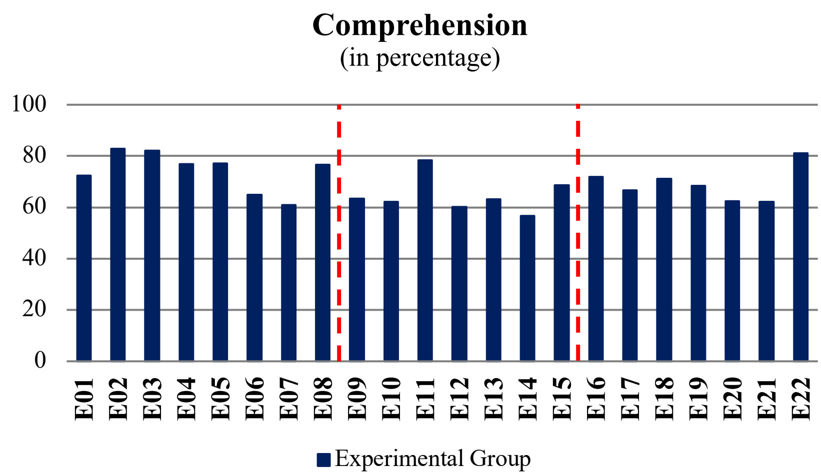


Figure 7.22 - Comprehension scores of Grade 10 participants throughout the academic year.

Regarding the comparison of mean content comprehension scores across the three terms, a Friedman test (N=23) was performed due to the lack of a normal distribution of the data in two of the terms. It revealed that such scores significantly varied across the school year ($\chi^2(2) = 8.96$, $p = .011$). Inspection of the mean values showed an initial decrease in the comprehension scores from T1 ($M = 74.22$) to T2 ($M = 64.67$), followed by a substantial increase from T2 to T3 ($M = 69.13$). These tendencies were later confirmed by Wilcoxon signed-rank tests comparing the percentage means of the three terms, which indicated that both the decrease and the increase between terms were statistically significant: T1 vs. T2 → $Z = -3.670$, $p = .000$; T1 vs. T3 → $Z = -2.650$, $p = .008$; T2 vs. T3 → $Z = -2.516$, $p = .012$.

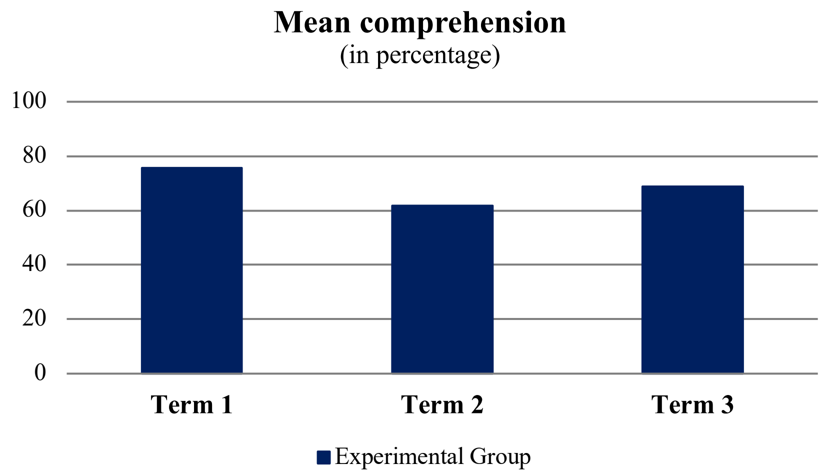


Figure 7.23 - Mean comprehension scores of Grade 10 participants across the three academic terms.

As the number of Grade 10 participants varied depending on the term and there was a change in the TV series that students watched (while in Term 1 they watched *I Love Lucy*, during the second and the third terms, they watched *Seinfeld*), data was also analysed term-independently.

In Term 1, 26 participants were selected to analyse the effect of time on the comprehension scores (those who had taken seven or eight comprehension tests). The mean comprehension score in Term 1 was 74.22%, which is the equivalent to 11.88 points. Besides, the lowest score in E07 represented 60.83% of comprehension, whereas the highest score in E02 represented 82.76% (see Table 7.24 for the descriptive results). A Friedman test showed a significant effect for time, indicating that scores significantly varied throughout the term ($\chi^2(7) = 51.065$, p = .000). Pairwise comparison further revealed that scores on E06 were significantly lower than those for E02 (p = .001) and E03 (p = .000). Besides, scores for E07 were also significantly lower than those for E02 (p = .000), E03 (p = .000), E04 (p = .008), E05 (p = .007) and E08 (p = .015).

Table 7.24
Comprehension tests' descriptive statistics for Grade 10 students at T1.

	I Love Lucy (in percentage)							
	E01	**E02**	**E03**	**E04**	**E05**	**E06**	**E07**	**E08**
N	30	29	29	30	29	30	30	29
Mean	72.50	82.76	82.11	76.86	77.16	64.79	60.83	76.72
SD	16.13	12.01	14.24	17.99	15.87	15.53	22.32	16.61
Min.	31.25	62.50	50	43.75	25	31.25	18.75	50
Max.	93.75	100	100	100	100	87.50	93.75	100

In Term 2, 23 participants were selected for the RM ANOVA to analyse the effects of time. Overall, scores seem to be a bit lower than in Term 1, as seen by the mean comprehension score, set at 64.67% (10.35 points), with the highest score being obtained in E11 (78.45%) and the lowest in E15 (68.75%). The RM ANOVA revealed a significant effect for time (F(6, 17) = 19.637, p = .000), which also showed that the effect size was large. Bonferroni post-hoc tests showed that there was a significant difference between E11 and the rest of the comprehension tests took in Term 2 (except for E15) (ps ranging from .000 to .011). Besides, pairwise comparison also revealed that scores for E15 were significantly higher than those obtained in E14 (p = .007) (see Table 7.25 for the descriptive results).

Table 7.25
Comprehension tests' descriptive statistics for Grade 10 students at T2.

	Seinfeld (in percentage)						
	E09	**E10**	**E11**	**E12**	**E13**	**E14**	**E15**
N	28	28	29	27	27	29	29
Mean	63.39	62.05	78.45	60.19	63.19	56.68	68.75
SD	16.99	21.04	16.59	20	21.25	16.61	20.48
Min.	25	18.75	37.50	25	31.25	25	18.75
Max.	93.75	100	100	100	100	93.75	100

Finally, in Term 3, 19 participants were selected for the longitudinal analysis. The mean comprehension score was 69.13% (11.06 points), in between the scores obtained in Terms 1 and 2. The episodes' comprehension scores varied between 62.27% (E21) and 81.01% (E22) (see Table 7.26). A Friedman test revealed that there was a statistically significant difference in the comprehension scores ($\chi^2(6) = 29.565$, $p = .000$), and pairwise comparisons further showed that the scores for E22 were significantly higher than those for E17 ($p = .003$), E19 ($p = .026$), E20 ($p = .002$) and E21 ($p = .000$).

Table 7.26
Comprehension tests' descriptive statistics for Grade 10 students at T3.

	Seinfeld (in percentage)						
	E16	**E17**	**E18**	**E19**	**E20**	**E21**	**E22**
N	26	26	26	24	26	27	26
Mean	71.88	66.59	71.15	68.49	62.50	62.27	81.01
SD	18.65	19.44	20.47	16.33	15	13.70	18.58
Min.	31.25	25	12.50	37.50	18.75	37.50	25
Max.	100	93.75	100	93.75	100	93.75	100

7.4.3. Content comprehension at Grade 6.

Figure 7.24 shows sixth graders' scores throughout the academic year. The two dotted lines mark the end of one term and the beginning of the next one. Additionally, the second dotted line also marks a change in the TV series (from *The Suite Life of Zack and Cody* to *Wizards of Waverly Place*).

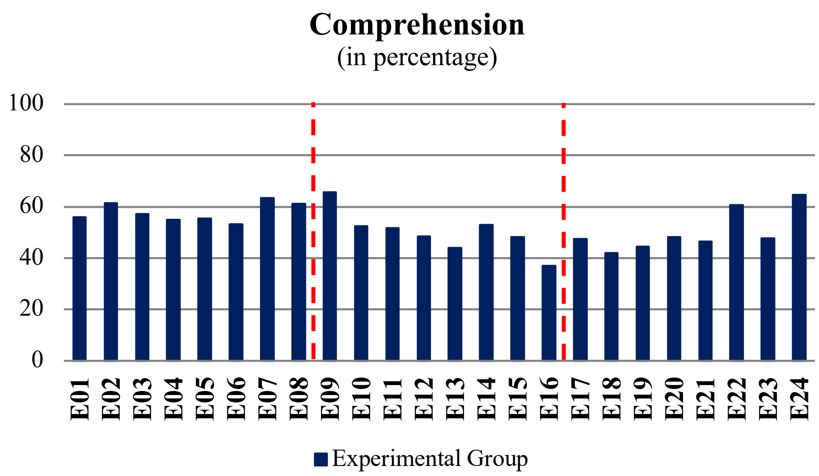


Figure 7.24 - Comprehension scores of Grade 6 participants throughout the academic year.

A one-way RM ANOVA with 18 learners revealed that mean comprehension scores were significantly different depending on the academic term ($F(2, 16) = 12.926$, $p = .000$). Besides, pairwise comparisons indicated that the mean comprehension score for Term 1 ($M = 57.84$) was significantly higher than that of Term 2 ($M = 50.07$) ($p = .000$) and Term 3 ($M = 50.27$) ($p = .036$). The mean comprehension scores for Terms 2 and 3 did not significantly vary ($p = .627$). Figure 7.25 on the next page shows the mean comprehension scores across the three academic terms.

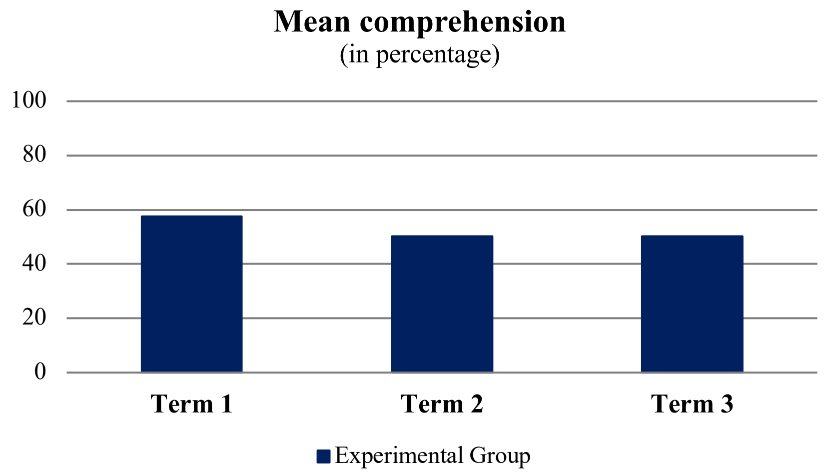


Figure 7.25 - Mean comprehension scores of Grade 6 participants across the three academic terms.

The number of participants allocated to the EG also varied across the year in Grade 6 ($N = 22$ in T1 and T3 and $N = 23$ in T2). At this level, learners watched *The Suite Life of Zack and Cody* in the first two terms and *Wizards of Waverly Place* in T3. That is why data were reanalysed term-independently.

Table 7.27 shows that comprehension in Term 1 ranged from 53.27% (E06) to 63.35% (E07) and the mean comprehension score of the eight episodes was 57.84% (9.25 points). A RM ANOVA (N=19) showed that time was not a significant factor in content comprehension throughout Term 1 ($F(7, 12) = 1.618$, $p = .221$). Besides, pairwise comparisons did not reveal any significant differences between any comprehension tests.

Table 7.27
Comprehension tests' descriptive statistics for Grade 6 students at T1.

	The Suite Life of Zack and Cody (in percentage)							
	E01	**E02**	**E03**	**E04**	**E05**	**E06**	**E07**	**E08**
N	21	22	22	21	22	21	22	22
Mean	55.95	61.36	57.10	55.06	55.56	53.27	63.35	61.08
SD	16.24	17.95	20.35	15.39	16.67	19.23	14.73	21.30
Min.	31.25	21.25	25	25	25	18.75	31.25	12.50
Max.	75	100	87.50	75	93.75	93.75	93.75	93.75

Table 7.28 shows that, in Term 2, when students continued watching *The Suite Life of Zack and Cody*, mean comprehension was situated at 50.07% (8.01 points), and that scores ranged from 36.97% (E16) to 65.76% (E09). A RM ANOVA (N=20) revealed a significant main effect for time ($F(7, 13) = 20.935$, $p = .000$) and Bonferroni post-hoc comparisons showed that the scores obtained in E09 were significantly different from the rest of the episodes, except for E14 (ps ranging from .000 to .032). In addition, scores for E14 were significantly higher than those obtained in E16 ($p = .046$).

Table 7.28
Comprehension tests' descriptive statistics for Grade 6 students at T2.

	The Suite Life of Zack and Cody (in percentage)							
	E09	**E10**	**E11**	**E12**	**E13**	**E14**	**E15**	**E16**
N	23	23	23	23	21	23	22	23
Mean	65.76	52.45	51.63	48.37	44.05	52.99	48.30	36.97
SD	11.13	17.24	21.59	15.10	17.62	20.97	19.50	14.95
Min.	37.50	25	18.75	25	6.25	18.75	18.75	18.75
Max.	87.50	93.75	87.50	81.25	75	100	81.25	68.75

In the third and final term, the TV series was new for all participants and these obtained a mean comprehension score of 50.27% (8.04 points), as seen in Table 7.29. The episode in which the highest score was achieved was E24 (64.77%) while participants obtained the lowest score in E18 (41.96%). The RM ANOVA (N=21) revealed a significant effect for time ($F(7, 14) = 3.437$, $p = .024$), indicating that the comprehension scores differed significantly depending on the episode. Bonferroni post-hoc comparisons revealed that the scores for E22 and E24 were significantly higher than in many of the previous episodes (E17, E18, E20 and E23), with significance values ranging from $p = .010$ to $p = .049$.

Table 7.29
Comprehension tests' descriptive statistics for Grade 6 students at T3.

	Wizards of Waverly Place (in percentage)							
	E17	**E18**	**E19**	**E20**	**E21**	**E22**	**E23**	**E24**
N	22	21	22	22	22	22	22	22
Mean	47.44	41.96	44.60	48.30	46.59	60.80	47.73	64.77
SD	21.88	16.32	20.16	18.01	18.47	21.23	17.52	25.04
Min.	12.50	12.50	6.25	25	18.75	12.50	18.75	18.75
Max.	100	87.50	81.25	87.50	87.50	100	87.50	100

7.4.4. Summary of the results for RQ2.

Overall, content comprehension of subtitled TV series proved to be participant- and episode-dependent, as no clear pattern was observed in the evolution of the scores on the comprehension tests. Time was found to have a significant effect at university, in the three terms in Grade 10 and in T2 and T3 in Grade 6. Pairwise comparisons also showed that significant differences between episodes did not follow a regular pattern, and that the first and the last episode of each term did not significantly differ in most cases. When content comprehension was studied throughout one academic year, it was seen that, in high school, the highest comprehension score was obtained in Term 1, and it significantly decreased in T2 to significantly increase again towards the end of the year. Sixth graders' comprehension scores significantly decreased from Term 1 to Terms 2 and 3, although they were very similar in these last two terms. For further information regarding specific scores on detailed, general and inferential comprehension questions, please check Appendix J.

7.5. RQ3 - Learning from TV series across proficiency levels

In the three preceding sections, vocabulary acquisition and retention from subtitled TV series as well as participants' degree of content comprehension have been thoroughly analysed for each proficiency group. However, up to now, no comparisons between proficiency groups have been drawn. This is precisely what the third RQ investigates ('How are L2 vocabulary learning and content comprehension affected by learners' proficiency level?') (see Table 7.30 for the summary of the results for this RQ).

7.5.1. Grade 6, Grade 10 and university.

Regarding relative gains for form, the results of a one-way between-groups ANOVA showed that there was a statistically significant difference at the $p < .05$ level ($F(2, 86) = 5.363$, $p = .006$). The actual difference between groups was quite large, as shown by the effect size, calculated using eta squared ($r = .11$). Post-hoc comparisons using the Tukey HSD test indicated that the mean score for university participants ($M = 35.23$) was significantly higher than sixth graders' ($M = 20.82$) ($p = .004$), while tenth graders' score ($M = 29.75$) did not significantly differ from either university ($p = .362$) or Grade 6 ($p = .133$) learners.

In relation to word meanings, another one-way between-groups ANOVA showed that there were statistically significant differences in the number of TWs' meanings learned by the three proficiency groups ($F(2, 86) = 8.255$, $p = .001$), and the effect size calculated using the eta squared ($r = .16$) showed that that the proficiency factor was very determinant in the learning of word meanings. Post-hoc comparisons using the Tukey HSD test showed that the scores obtained by the university group ($M = 23.26$) were significantly higher than those achieved by Grade 6 participants ($M = 10.33$) ($p = .000$), while tenth graders' scores ($M = 16.39$) did not significantly differ from either university ($p = .058$) or Grade 6 ($p = .178$) learners. Hence, the only significant difference found between proficiency groups as far as vocabulary learning is concerned (regardless of the lexical aspect) was between university and Grade 6 learners.

Table 7.30
Summary of the findings for RQ3.

Term	Proficiency group	Vocabulary		Content comprehension
		Form	Meaning	
Term 1	G6 vs. **G10**	✗ ($p = .133$)	✗ ($p = .178$)	✓ ($p = .000$)
	G6 vs. **Uni**	✓ ($p = .004$)	✓ ($p = .000$)	✓ ($p = .000$)
	G10 vs. Uni	✗ ($p = .362$)	✗ ($p = .058$)	✗ ($p = .081$)
Term 2	G6 vs. **G10**	✓ ($p = .000$)	✗ ($p = .645$)	✓ ($p = .005$)
Term 3	G6 vs. **G10**	✓ ($p = .000$)	✓ ($p = .006$)	✓ ($p = .000$)

Note. Significant differences are always in favour of the proficiency group highlighted in **bold**.

Focusing on content comprehension, a Kruskal-Wallis test was performed because data was not normally distributed in the university group. It revealed statistically significant differences across the three proficiency groups ($\chi^2(2) = 35.756$, $p = .000$). Independent samples t-tests and

Mann-Whitney *U* tests between pairs of groups showed that the comprehension scores obtained by Grade 6 participants (M = 57.47) were significantly lower than tenth graders' (M = 75.72) ($t(50)$ = -5.381, p = .000). Similarly, sixth graders' scores were also significantly lower than university participants' (U = 39, z = -5.816, p = .000). Finally, another Mann-Whitney *U* test revealed that tenth graders' and university participants' scores did not significantly differed from one another (U = 429, z = -1.742, p = .081).

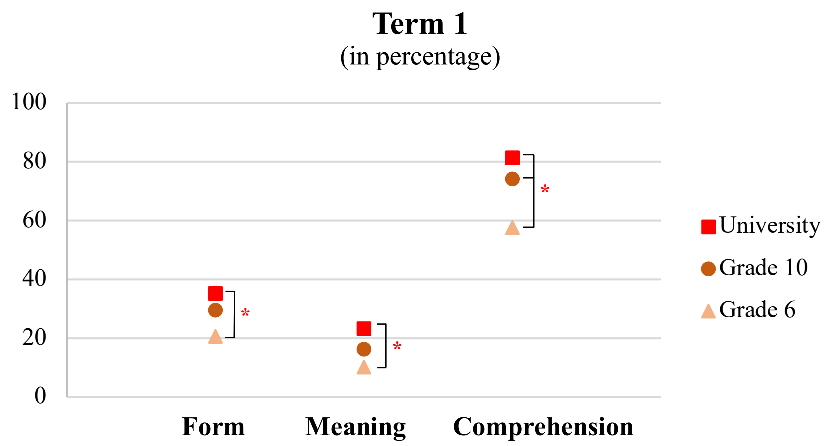


Figure 7.26 - Vocabulary and comprehension scores across proficiency groups in Term 1. | *Note.* Significant differences are marked with an **asterisk**.

7.5.2. Grade 6 and Grade 10.

Grades 6 and 10 were compared in Terms 2 and 3. In the first of these two terms (T2), an independent samples *t*-test showed that relative gains for form obtained by Grade 10 students (M = 33.55) were significantly superior to those attained in Grade 6 (M = 18.08) ($t(47.932)$ = -4.105, p = .000). The magnitude of the differences in the means was very large (r = .25). However, as far as word meanings are concerned, a Mann-Whitney *U* test was run and showed that, although Grade 10 students (M = 16.39) obtained higher mean values than Grade 6 learners (M = 10.33), the difference did not reach statistical significance (U = 308.5, z = -.461, p = .645).

Focusing on content comprehension, another Mann-Whitney *U* test indicated that the scores obtained by Grade 10 students ($M = 64.67$) were significantly higher than sixth graders' records ($M = 50.07$) ($U = 174$, $z = -2.802$, $p = .005$).

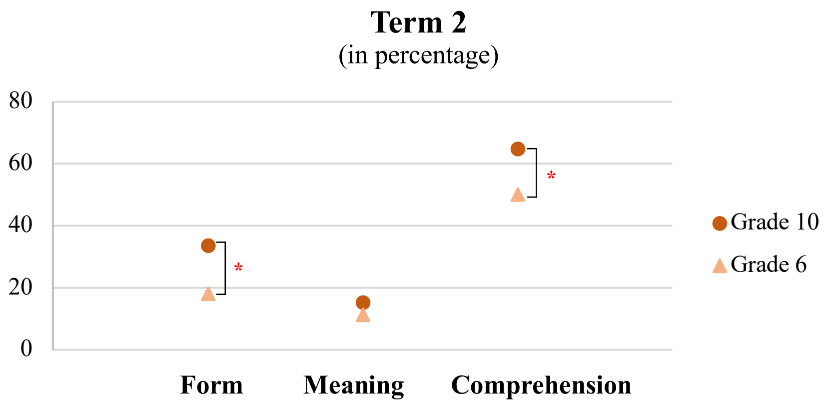


Figure 7.27 - Vocabulary and comprehension scores across proficiency groups in Term 2. | *Note.* Significant differences are marked with an **asterisk**.

Finally, in the last term, an independent samples *t*-test showed that there were significant differences in the number of TWs' forms learned by Grade 6 ($M = 20.19$) and Grade 10 ($M = 38.64$) EFL learners ($t(44.849) = -4.186$, $p = .000$). The difference in the means was large, as indicated by the eta squared ($r = .27$). There were also significant differences in relative gains for meaning (Grade 6 - $M = 11.70$ vs. Grade 10 - $M = 25.11$), as shown by a Mann-Whitney *U* test ($U = 167.5$, $z = -2.747$, $p = .006$). The test also revealed that the effect size was very large ($r = .39$).

Regarding content comprehension of the TV series, another Mann-Whitney *U* test showed significant differences between the two groups ($U = 91$, $z = -4.142$, $p = .000$), further proving that the comprehension index obtained in Grade 10 ($M = 69.13$) was significantly higher than that observed in Grade 6 ($M = 50.27$).

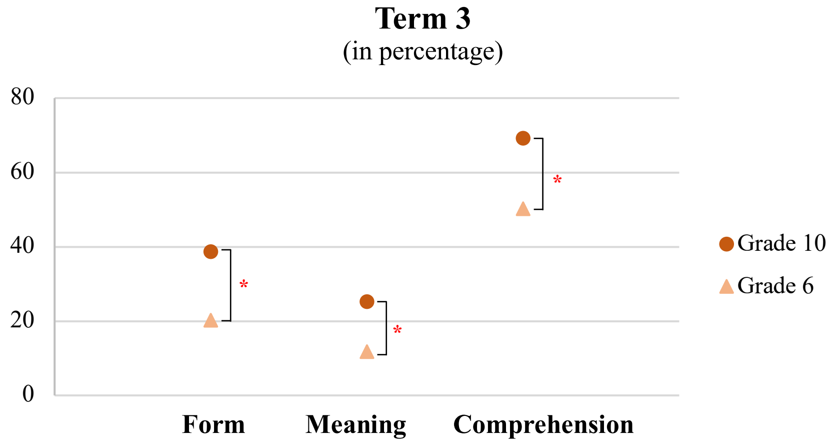


Figure 7.28 - Vocabulary and comprehension scores across proficiency groups in Term 3. | *Note.* Significant differences are marked with an **asterisk**.

7.5.3. Summary of the results for RQ3.

The vocabulary scores (both word form and word meaning) obtained by the university group were significantly higher than sixth graders' scores, but not than tenth graders'. Regarding the comparison between Grade 6 and Grade 10, the two groups' vocabulary learning scores did not differ in Term 1, but were significantly different in Term 2 (only for word meaning) and in Term 3 (for both word form and word meaning), always in favour of Grade 10 learners.

As far as content comprehension is concerned, university participants significantly outperformed sixth graders, but not high-school students. Finally, Grade 10 learners significantly outdid Grade 6 participants in all the three terms the pedagogical intervention was divided into.

CHAPTER EIGHT - DISCUSSION

This chapter will be devoted to the discussion of the results presented in Chapter 7. These will be considered following the RQs that this dissertation proposed and grouped into different topics to help to relate our findings to previous research and to our theoretical framework. First, results on vocabulary learning from viewing TV series will be discussed, followed by the interpretation of the results of the vocabulary delayed test. Then, our focus will be on content comprehension of the TV series. Special attention will be put in the role of proficiency and the extent to which it may affect lexical acquisition and comprehension in extensive TV viewing.

8.1. Learning of foreign language vocabulary from subtitled TV series

The first RQ enquired about vocabulary learning from extended exposure to TV series at three proficiency levels (university, high school and primary school). At each level, half of the participants were allocated to the CG, who followed regular vocabulary instruction through a set of learning tasks, and the remaining half to the EG, who was additionally exposed to subtitled TV series for an academic term or year. Relative vocabulary gains for word form and word meaning were then calculated and compared across experimental conditions.

The results of the first RQ showed that vocabulary acquisition from watching subtitled TV series took place at the three proficiency levels analysed. At almost all testing times, the EGs tend to gain more than the CGs. However, significant differences between EG and CG were not found at university. In contrast, in high school, there were significant differences between groups throughout the academic year, for both word form and word meaning learning. Finally, in primary school, they were only found in Term 3, towards the end of the pedagogical intervention. Results also showed that proficiency level affected the vocabulary gains made by the participants watching the TV series: more proficient learners tended to obtain significantly greater vocabulary gains than lower-proficient students.

8.1.1. Multimedia and cognitive load theories: L2 vocabulary learning.

As seen in Chapter 2, this doctoral dissertation is framed by two main theories supporting that multimodal input is beneficial for language learning: the DCT (Paivio, 1986), according to which concepts presented through both the verbal and the non-verbal channel are better learned and recalled; and the CTML (Mayer, 2009), which claims that people learn better from words and picture than from words alone. The thesis also draws on the CLT (Sweller, 1994), which defends that our brain has a limited cognitive capacity that should not be overloaded.

Results from the university group do not seem to support the DCT. As opposed to what Sadoski (2015) claimed, the two-format presentation, by which content is given concrete referents and explanatory language is further supported by non-verbal information, was not conducive to significantly better learning. In other words, the extra help in the connection between the verbal and pictorial modes in the form of multimodal input did not lead to larger gains. Presumably, there could have been much unnecessary information for the EG. First, it is possible that explicit teaching was enough at this level to learn (some of) the TWs. It is also possible that English captions at this level presented information that learners already obtained from listening to the audio and viewing the video; although captions may have been used to confirm or disconfirm whether or not learners actually understood the TV series. In such cases, captions could have presented redundant information although students could automatically read them even if they did not need to (d'Ydewalle & De Bruycker, 2007). This could have distracted the participants' attention, leaving less time and processing resources available for vocabulary learning. In Grade 10, though, the results do support the DCT and it is reasonable to think that stimulation of the verbal and imagery subsystems can have effects on depth of processing and lead to better recall. The verbal subsystem was stimulated by the presence of the aural (audio) and written (subtitles) text and the imagery subsystem was activated thanks to the video participants in the EG were additionally exposed to. The activation of referential connections (the link between imagens and logogens) was possible in the EG but not in the control condition, since only logogens (as representations of all language phenomena) were available to learners of the CG. Thus, this double processing in the EG at the intermediate level could have helped learners to make the most of the intervention.

In the case of beginner learners, it was just after several viewing sessions that videos with L1 subtitles backed up the main tenets of the DCT. Once EG learners were used to watching TV in English, it might have been easier for them to merge the three modes of input they were exposed to (video, audio and subtitles), simultaneously process them and consequently made the verbal and imagery subsystems work interconnectedly. This was not possible the first two terms, when the low level and the lack of practice could explain why the same results were obtained by learners in the two conditions. In this respect, in Terms 1 and 2, the verbal and non-verbal subsystems may have continued working parallelly and independently, so there was no connection between them and, as expected, this did not lead to greater learning.

The CTML can also contribute to explaining the differences in vocabulary acquisition between the two conditions in Grades 6 and 10. In light of the results, the multimedia principle (Mayer, 2009, 2014), by which people learn more deeply from words and pictures than from words alone, holds true throughout the academic year in Grade 10 and in the last term in Grade 6. The fact that all the TWs were presented in the vocabulary pre-task and that students were familiar with the TV series (after some viewing sessions) could contribute to the learning of the target vocabulary, as suggested by the pre-training principle. The temporal contiguity principle (the three modes of information were presented simultaneously, and not in successive presentations), and the fact that pictures were combined with spoken discourse in the video (i.e., following the modality principle) could have also facilitated the acquisition of unknown vocabulary when exposed to multimedia presentations.

In relation to the CLT, results at university do not suggest that the EG had its cognitive capacity overloaded due to the presence of the video, since the EG outperformed the CG, even if significant differences between groups were not found. Something similar happens with tenth graders, when participants seem to have enough cognitive and attention resources, distributed between the two subsystems –verbal and imagery–, to cope with the materials they were exposed to (significant differences were never found in favour of the CG). Additionally, textual support might have contributed to reduce CL; more specifically, L2 subtitles may have assisted in reducing *extraneous* CL and increasing the *germane* type, which might have led to more learning and less chances of being overloaded. The presence of subtitles might have freed more cognitive resources that participants could allocate to learning the target vocabulary and to make the most out of the video viewing experience. However, in Grade 6 participants, the *intrinsic* CL of the video-viewing task may be too high for beginner learners, and adding the subtitles did not lower *extraneous* CL, but probably increased it (Kruger et al., 2013; Mayer et al., 2014). This consequently may have exceeded the limits of cognitive processing and WM capacity (Baddeley, 1992), to prevent more significant language acquisition. However, although TV series viewing did not contribute much to the learning of new vocabulary, it is possible that subtitles could have assisted low-level learners in understanding the video input they were exposed to by providing an L1 translation of what was being said in the episodes.

Throughout the pedagogical intervention, CL could have also been reduced thanks to the use of worked examples (Kalyuga et al., 2001). In this respect, the pre-teaching of the TWs anticipated some of the content to learners and also increased their familiarity with the target items. In a way, it could be said that all the TWs, at the end of the term, were worked examples, since they had already been introduced at least four times (pre-task, video (x2) and post-task) and tested once (pre-test). This freed some learning resources and reduced learners' anxiety as they knew which content was supposed to be tested. Besides, the same test format was used throughout the year, so less attention had to be directed towards task completion instructions, thus reducing the cognitive demands of the task (Darmi, 2012).

In sum, it seems that not all three theories are confirmed to the same extent based on the results of this dissertation. The DCT and the CTML seem to support the results at intermediate levels, and to a lesser extent at beginner levels, where the cognitive overload could have been too high. The alleged benefits for language learning of the DCT and the CTML do not seem to be applicable at university, at least to the sample that was analysed in the present thesis. However, it should be taken into account that the intervention at this proficiency level lasted for one academic term (eight viewing sessions).

8.1.2. The role of proficiency in vocabulary learning.

In relation to the role that proficiency plays on the learning of vocabulary through exposure to multimodal input, no studies that we know of have compared three different proficiencies following a longitudinal research design with intergroup analyses. The design of the thesis allows to compare groups at different proficiency levels (beginner, intermediate and upper-intermediate) and see the effects of TL competence on the acquisition of vocabulary from video exposure. Most of previous research on the topic relied on one group of participants and explored whether scores on VS tests or general proficiency measures for that group explained vocabulary gains. In the few studies where two groups of learners were available, authors typically compared learners close in age and at the beginning of the language learning process (e.g., Grade 4 vs. Grade 6; Koolstra & Beentjes, 1999). As can be seen, this is inherently different from the present dissertation.

In light of the results of this thesis, we can claim that, when significant differences in vocabulary learning are found between EGs, it is the more proficient group who benefited more from TV series viewing, as Grade 6 students obtained significantly lower gains than Grade 10 and university learners. University participants also scored higher than high-school students (although no significant differences were found). In this respect, we cannot leave aside the role that prior vocabulary knowledge plays on lexical acquisition: in the last few years, some studies have also related VS to multimodal input (e.g., Alexiou, 2015; Montero Perez et al., 2013; Peters et al., 2016; Peters & Webb, 2018) and found that having a larger vocabulary enhances lexical acquisition. With the exception of some studies (e.g., Frumuselu et al., 2015), the majority of research found traces of the Matthew effect (Stanovich, 1986), or the rich-get-richer principle, according to which more proficient learners gain more than less skilled learners. Other studies (e.g., d'Ydewalle & Van de Poel, 1999; Koolstra & Beentjes, 1999) also found that older students learned more vocabulary than younger ones. Therefore, the present results would mostly support and corroborate previous research, despite university students not scoring significantly higher than Grade 10 learners (the smaller gap in proficiency between these two groups in comparison to Grade 6 may account for the results).

Another possible explanation for the decisive role of proficiency could be connected to metacognition, or "thinking about thinking" (Anderson, 2002, p. 3). Intermediate participants learned how to learn and developed metacognitive strategies, which were not yet developed by beginner learners, and that allowed them to benefit more from the intervention. In this sense, "learners who are metacognitively aware know what to do when they do not know what to do" and this is conducive to "more profound learning and improved performance" (Anderson, 2002, p. 3). Beginner learners, in contrast, had a more limited metacognition, and they probably did little monitoring of their own learning (Flavell, 1979). Furthermore, as higher metacognitive indices have been linked to more vocabulary learning (Rasekh & Ranjbary, 2003), intermediate participants could benefit from this advantage and learn more target items whereas lower-level learners could not.

This is in line with the development of learner autonomy, or the "ability to take charge of one's own learning" (Holec, 1981, p. 3), which research has shown to be directly and positively linked to general proficiency (Dafei, 2007). High-school and university participants were more autonomous learners and did not rely so much on guided and directed teaching and were able to go beyond what the teacher or researcher expounded. In contrast, primary school children had not yet fully developed this ability and they relied more on the 'expert', so they were less able to 'explore' language on their own.

Moving on to the between-group comparisons, results showed that university and Grade 10 participants significantly outperformed sixth graders. It is reasonable to say that these differences were expected. There was a four- and seven-year difference between the two groups and university students had had, at least, more than 400 hours of formal instruction, plus many more chances to get in contact with the TL. For instance, they had participated in summer camps or stays abroad where English was spoken and had attended more extracurricular classes. Besides, their familiarity with FL video viewing was also greater, as some reported viewing L2 television in their free time. Leaving this aside, there are also vast differences in processing abilities and rote learning. As has been pointed out, the significant differences can be due to sixth graders' limited cognitive capacity, which impedes them from fully benefitting from the intervention. In line with this, the viewing task was challenging for beginner learners and, in contrast, it was quite familiar for secondary school and university students.

However, it also needs to be pointed out that differences between Grade 10 and university participants were not significant, always bearing in mind that the intervention only lasted one academic term at university. University freshman students were just three years older than high-school students and had received 200 more hours of curricular instruction, which, for FL development is not actually a lot of input, especially in a community where the TL is not spoken. Besides, these two groups are supposed to have similar attention resources and related reading and aural skills. This situation would be comparable to Neuman and Kosninen (1992), who divided their sample into three proficiency levels according to the participants' fluency in the TL (limited, fluent and mastery) and did not find any differences between the two more proficient groups.

Focusing on each of the three levels, general proficiency may also explain the lack of statistically significant differences between conditions observed in the university group. Upper-intermediate university participants were proficient enough to learn the target items when they were pre-taught at the beginning of each viewing session. The fact that the researcher or the teacher corrected the vocabulary pre-task immediately after its administration and clarified all the doubts participants had regarding the target vocabulary was a good learning opportunity for them. Besides, their hypotheses could be later confirmed or rejected on the vocabulary post-task, at the end of each viewing session, when they encountered the TWs' forms and meanings again. In other words, although one of the groups was exposed to video input and received more exposure to the TWs, the learning observed in both groups was probably due to deliberate teaching. That is, the input received at the beginning of each viewing session was sufficient for all learners to acquire the TWs; adding extra exposure through multimodal input may have possibly consolidated this knowledge or provided instances of use of the TWs, but that was not measured in the post-test. Hence, it should also be admitted that, at this proficiency level, formal teacher-led instruction was already effective to learn the target vocabulary up to the same extent.

Although differences between conditions were not found in the university group, they appeared in Grade 10, a group with a more limited proficiency level. This first demonstrates that learning of vocabulary does occur through watching television (e.g., Baltova, 1999; Sydorenko, 2010; Winke et al., 2010) and that this practice is conducive to greater learning when combined with more traditional forms of instruction, at least at this proficiency level. A possible explanation for these benefits is that, thanks to the video with English subtitles, the EG could make more sense of the input they were exposed to. In other words, high-schoolers may not have had enough with the formal instruction at the beginning and end of each session, which may have partially helped in learning some of the target vocabulary, but video viewing enhanced learning. Students benefitted from the presence of visual and textual support since their reading and aural skills are not yet fully developed, and assisted them in following the audiovisual materials they were being exposed to.

Continuing with a similar argument, greater attention capacity and higher proficiency level are normally positively correlated (Tse & Altarriba, 2014). Thus, intermediate learners' attention resources are more limited than more proficient learners' (e.g., university) and they need to be split among the different tasks that language learners are supposed to do. In the context of the pedagogical intervention, these needed to be allocated to several tasks: (1) understanding the English audio, (2) reading the English captions, and (3) paying attention to the video and the storyline of the episodes. Consequently, learners at the intermediate level may have really valued the presence of the textual support so as to facilitate the understanding of the video and have more free resources to devote to the vocabulary learning task. If video and subtitles had not been available, much of the participants' attention could have been exclusively devoted to speech segmentation (Mitterer & McQueen, 2009), hindering vocabulary acquisition and content comprehension. At the same time, explicit instruction without further encounters with the TWs (Schmitt & Schmitt, 1995; Webb, 2007) was not enough for the CG to catch up with the EG.

Turning to Grade 6 results, significant differences between control and EGs only appeared in the third term. To start with, the first conclusion we can draw from the intervention at this level is that vocabulary acquisition can take place when low-level learners are exposed to FL television, even if such acquisition is very limited (Alexiou, 2015; Rice et al., 1990). However, for this to happen, a lot of input, training and familiarity with the task and the audiovisual materials may be needed. Grade 6 learners had a very limited English proficiency, with only 900 hours of formal instruction and six years of learning prior to the beginning of the intervention. It is presumable that learners at this age and level would need more input and extra help to acquire some of the TWs, through both traditional classroom instruction and video viewing. For this reason, L1 subtitles were used, as opposed to bimodal subtitles, and the vocabulary pre-tasks included more images, matching exercises and a higher presence of L1 translations.

In connection to the beginners' proficiency level, the materials the EG was exposed to could have been too challenging for them, especially because their English VS was very limited (1,750 words) and the average vocabulary demands needed to adequately understand the television episodes were around the first three thousand words of the TL. Actually, Webb and Rodgers (2009b) showed that knowledge of the most frequent 3,000 WFs (plus proper nouns and marginal words) were needed to reach 95% coverage of American and British TV programmes, so most Grade 6 students could fall short of VS even if the TV series chosen targeted preadolescents and the subtitles were always in the L1. In consequence, the input could have been too demanding for them (even incomprehensible sometimes) and this may have discouraged them from paying close attention to the vocabulary appearing in the TV series and could have diminished video's potential benefits. Additionally, subtitles could have been a distractor rather than a facilitator, especially because materials may not have been adapted to the learners' level (Danan, 2004): they could have been useful to follow the plot of the episodes, but not to learn the target vocabulary. Learners at this age and proficiency level could have been overwhelmed by the presence of subtitles, and chances are that they did not know how to cope with them. In this sense, even if subtitles were in the students' L1, they may have been displayed very quickly for these learners (Gambier, 2015), as their L1 reading fluency was still developing (Landerl & Wimmer, 2008).

Overall, the results seem to support Vanderplank's threshold hypothesis, according to which learners need to attain a certain proficiency level to start benefitting from captioning and subtitling. It could be claimed that learners would need to have an intermediate level of the TL to profit from the additional exposure to video viewing. Below this level, benefits are not clear, at least in the short term, and they may only arise with sustained exposure to multimodal input. Above the intermediate level, multimodal input's positive effects for the learning of the target vocabulary were not detected, at least in comparison to formal regular instruction. However, the intervention comprised just eight sessions.

With learners below the intermediate level, limited L1 / L2 reading and L2 listening skills can impede them to successfully benefit from subtitled video viewing. Furthermore, results make us think that, when original non-adapted materials are used at beginner levels, the cognitive and attention demands of watching FL television are too high and they can balance out the benefits that viewing may have at higher levels. If learners are above the intermediate level, chances are that curricular explicit instruction will be as effective as video viewing: it is also less time consuming and exposure to multimodal input will not add much to the vocabulary learned from formal teaching (possibly, unless exposure is massive). Based on our findings, the gains when watching subtitled TV series are more clearly seen in intermediate-level classes, although explicit instruction with focus-on-forms activities to foster the learning of the target vocabulary is also recommended.

8.1.3. Sustained exposure in vocabulary learning.

At the proficiency levels where the intervention lasted one academic year (i.e., Grade 6 and Grade 10), benefits in the experimental condition are more obvious in Term 3. This could be explained by the task repetition and familiarity effect (Bygate & Samuda, 2005) and its effects on vocabulary use (Kim, Crossley, Jung, Kyle, & Kang, 2018). After several months of being exposed to subtitled TV series, learners may have developed some strategies which helped them to cope with and benefit from the audiovisual input. For instance, the participants in Grade 10 and Grade 6 could have paid more attention to the target items encountered in the vocabulary pre-task and tried to spot them while viewing the episode.

At the end of the first term, sixth and tenth graders had only been exposed to eight episodes, lasting around three hours in total. Even if it is more than what they usually watch at school, it is derisory if compared to the average leisure time spent watching TV at home. For instance, in 2017, 88% of Spanish citizens watched TV daily (European Commission, 2017) and they spent an average of 215 minutes each day doing this activity. Hence, it took eight weeks for participants in the EGs to receive the same input that Spaniards are exposed to in less than one day. It is then reasonable to think that they actually needed much more input to start benefitting from the experiment. The same holds true for the second term because multimodal input failed to show clear and consistent significant differences. At the end of this term, learners had already received five and a half hours of multimodal input, the same as watching TV for one and a half days in the average Spanish home. In contrast, though, at the end of the year, learners had been exposed to more than eight hours of multimodal input. Hence, the gap between the amount of input received by the two experimental conditions was proportionally and increasingly larger and so the benefits of the pedagogical intervention started to be seen. It could be said that sustained exposure to multimodal input was more beneficial than short-term exposure, which is further shown by the results of the mixed ANOVAs, which revealed a significant main effect for time in both grades as far as the acquisition of vocabulary is concerned.

An additional point that needs to be taken into consideration is the fact that longitudinal exposure to multimodal input was beneficial regardless of the changes in the TV series that took place throughout the year. While in Grade 10 the TV series was replaced by a new one at the end of Term 1, in Grade 6 this did not happen until the end of Term 2. Hence, at different points of the intervention, tenth and sixth graders had to get familiar with new characters and their relationships, the setting and the recurrent punchlines of the episodes, but this did not have an impact on the learning of the target vocabulary. It was in Term 3 when significant differences in both grades were found, even if, for example, beginner learners had just begun to watch *Wizards of Waverly Place* after two terms watching *The Suite Life of Zack and Cody*. This TV series was shown to be comparable in terms of difficulty to *The Suite Life of Zack and Cody*, as we have explained in the 'Instruments' section (see 6.5.1.1.).

In sum, we could conclude that sustained exposure is a crucial aspect that needs to be considered for video viewing to be most effective in language learning. Long-term effects could not be explored at university since, for logistic reasons, participants at this proficiency level could not be followed for more than one academic term. However, in Grades 6 and 10, differences favouring the experimental condition are to be found in a more consistent way towards the end of the academic year.

8.1.4. The acquisition of word forms and word meanings.

L2 subtitles are said to facilitate the acquisition of TWs in intermediate learners (Frumuselu et al., 2015; Peters et al., 2016; Zarei & Rashvand, 2011), and this thesis also finds that both word forms and meanings were better learned at this level. In Grade 10, in two of the terms (T2 and T3), significant differences were observed in the number of forms participants were able to recall. When participants are exposed to bimodal subtitles or captions (L2 audio + L2 text), as was the case here, they can read the written forms of the TWs on the screen and this helps them to distinguish between separate words (Danan, 2004; Peters et al., 2016), which aids speech segmentation. The fact that participants saw the correct written forms of the TWs would, theoretically, facilitate their learning (Caimi, 2006; Danan, 2004). However, this did not happen in Term 1, probably because of the novelty of the experience for learners.

While bimodal subtitles (L2 audio + L2 text) have been found to favour the acquisition of word forms, standard subtitles (L2 audio + L1 text) should favour the acquisition of the TWs' meanings (Koolstra & Beentjes, 1999), since the L1 translations are available for students to resort to them if needed. However, in our study, standard subtitles only helped to establish form-meaning connections in the last term, as shown by the results in Grade 6 (the only group exposed to this subtitling condition). Actually, in the only term when differences between groups were significant, both word forms and meanings were better learned by the EG. In line with Peters et al. (2016), who defend that linking the meaning of the target item to the L2 aural form in the speech stream might be difficult, L1 subtitles did lead to more of this linking in the last term, but failed to do so in Terms 1 and 2 probably due to the difficulty of establishing such connections for beginner learners.

Furthermore, so as to learn the meanings of the TWs, a certain degree of imagery is also needed. Rodgers (2018) claims that for learners to acquire an unknown word, it is better if its image "appears at a time in close proximity to the aural occurrence of the word" (p. 192). Possibly, *Wizards of Waverly Place*, the TV series used in Term 3, had a higher degree of imagery than *The Suite Life of Zack and Cody*; facilitating the learning of TW meanings. For instance, the meaning of the TW *wand* could be easily inferred in the TV series as, in many occasions, it is verbalized at the same time some of the characters use one to cast a spell. Hence, it could be that the degree of imagery of the TWs is more relevant when learning their meanings than the language of the subtitles, even if they are in the students' L1. However, prior to drawing such conclusion, an in-depth study needs to be conducted to see the extent to which the TWs were verbalised at the same time they appeared on screen, and to evaluate the effects of imagery on word learning in the materials used for the pedagogical intervention. This was not the purpose of the present investigation, though.

Yet, it should be noted that making the distinction between form and meaning allowed us to see that subtitles can help in initial stages of word recognition (independently of whether they are in L1 or L2 and of whether learners can or cannot actually attach a meaning to that particular form).

8.1.5. Interpreting vocabulary gains in relation to previous research.

First of all, it should be acknowledged that comparisons with previous research are very difficult to draw, since no studies that we are aware of have conducted this type of interventions in real FL classes.

Focusing on university learners, we found no differences between conditions. This differs from both Rodgers (2013) and Peters and Webb (2018), who had found a positive effect of TV viewing on vocabulary acquisition. However, it must be borne in mind that the CGs in these previous studies did not receive explicit instruction of the target vocabulary and were only exposed to it through the pre- and post-tests (incidental learning), which is different from what our CG did. In our study, learning was a focused process and students, after some sessions, knew what would be asked in the post-test. If the present intervention had tested incidental learning, with no pre-task, it is presumable that scores would have been lower (Schmitt, 2008). More differences between the present and previous research lie on the fact that the multimodal input selected for the study was of different nature: while Rodgers (2013) selected 10 episodes of a TV series with a mean running time of almost 43 minutes each, Peters and Webb (2018) chose a one-hour long documentary. Our university group was only exposed to episodes of 20 minutes, not even half the length of Rodgers' episodes. It is highly probable that our upper-intermediate students needed to accumulate larger amounts of multimodal input so as to start benefitting from the viewing practice. Moreover, the testing format adopted by Rodgers (a meaning recognition test) was very different to ours. Peters and Webb used a meaning recall test, similar to ours, but it was administered immediately after the viewing of the documentary, not at the end of the term, which makes the studies more difficult to compare.

In terms of learning gains, our university participants learned 35.22% –EG– and 33.14% –CG– of the TWs' forms while they gained 22.88% –EG– and 19.52% –CG- of the word meanings, which indicates that gains were higher than those found in the previous two studies. While Rodgers reported gains of 23% –tough test– and 29.6% –sensitive test– by the EG and 20.8% –tough test– and 25.4% –sensitive test– by the CG, Peters and Webb's (2018) participants gained, on average, 22.27% on the form recognition test, 5.83% on the meaning recall test and 9.95% on the meaning recognition test. Besides, in the test we used, decontextualized productive knowledge of the forms and meanings of the TWs was required, while the other studies mostly tested receptive abilities, which research has shown to be less challenging to master (Laufer, 1998; Webb, 2008); this is coherent if we think that there was explicit instruction in our case. We also need to bear in mind that our vocabulary post-test was administered at the end of each term, not immediately after each viewing session, which implies that some TWs were tested eight weeks after they were first encountered in the vocabulary pre-task, contributing to explain the indices of vocabulary learning.

Regarding the results obtained by intermediate learners, a population which previous research has shown to benefit the most from captioning (Montero Perez et al., 2013), our EG participants learned an average of 35.17% of TWs' forms and 20.08% of word meanings whereas the CG acquired 27.78% and 12.62%, respectively. These results are comparable with those observed by the two pieces of research commented before. They fall a bit below Rodgers (2013), but are above Peters and Webb (2018). The fact that Rodgers (2013) used a meaning recognition format, easier than our pre- / post-tests, could explain this difference, together with the fact that his students were older and with more language learning experience. As observed, the gains found in this study were considerably larger than in Peters and Webb (2018), possibly because of the long-term exposure to video viewing, as opposed to a one-hour TV programme. This seems to support the importance of repeated exposures and accumulation of input throughout the intervention.

In Grade 6, vocabulary learning from video viewing was possible but very limited (as in d'Ydewalle & Van de Poel, 1999 or Rice et al., 1990). Sixth graders in the experimental condition learned an average of 18.38% of the TWs' forms and 14.78% of word meanings, which is the same as saying that they learned 7.35 word forms and 5.91 word meanings. It is important to highlight that these figures were obtained after explicit focusing on the target items and doing vocabulary learning tasks. In consequence, it might not be erroneous to think that incidental vocabulary acquisition through multimodal input exposure would have been too challenging for Grade 6 learners: even with the help of L1 subtitles and vocabulary tasks, participants were able to learn a very limited number of target items.

Although comparisons with previous research are very difficult to draw due to the absence of similar studies, it is worth reminding others with beginner learners. For instance, Alexiou (2015) found that four- and six-year olds recognised a third of the 21 TWs which appeared on the videos they were shown. However, test type was very different (form recognition vs. meaning and form recall), the number of TWs was also smaller (21 in total vs. 40 each term) and the population even younger. The same applies to d'Ydewalle and Van de Poel (1999): their tests tapped on meaning recognition and learners were asked to select the best translation of the ten target items that had been preselected. Results showed they knew around 50% of the words when the test was conducted with Danish as the TL, whereas the mean correctness rate for the French version was lower. Both figures, however, are far above the 14.78% of TWs' meanings which were learned on average by our EG. This was somehow expected since participants in our study had to remember more and to a deeper extent. Besides, the number of target items in their study was also smaller and less cognitively demanding than being attentive to 24 episodes of a TV series and picking up the forms and meanings of five TWs each week. Furthermore, the results from the first two terms, when no significant differences were found between the two conditions, back up what Galimberti (2016) saw using the same TV series and in a very similar context (EFL classroom learning in Italy). It is worth pointing out that the two studies have a lot in common: similar pre- and post-tests; same TV series (she selected one episode of *The Suite Life of Zack and Cody*), and similar population (Grade 6 learners). Thus, the comparison is pertinent. First, the fact that 10% of the TWs were learned incidentally in her study supports the limited acquisition levels found in this doctoral thesis. Besides, the additional presence of textual support did not help much in vocabulary recalling, as no differences were found between the uncaptioned condition and the L1 and L2 subtitled conditions. This is in line with the results found in the present study, when the control condition obtained similar learning gains than the EG in two terms. In sum, we can see that the results of the two studies are comparable, bearing in mind that frequency and length of exposure to the input were substantially distinct and that she did not include a non-video condition.

Finally, the increase in the indices of vocabulary acquisition in Term 3 in Grades 6 and 10 is in line with Frumuselu's doctoral dissertation (mainly, studies 1 and 3), when the amount of vocabulary learned after seven weeks of exposure to TV series with bimodal subtitles (M=14.68) was poorer than the number of TWs learned after 14 weeks (M=16.14). Bravo (2008) also found a similar pattern and saw that Portuguese high-school learners of English performed better in the second half of a 10-week treatment when they watched ten episodes of a well-known TV series with bimodal or standard subtitles (M=76.9%, as opposed to M=59% in the first half). Besides, the results from the last session in this ten-week period almost doubled those obtained on the pre-test, at the beginning of the study. However, no statistical tests were run in her study to see if the increase was significant and to what extent.

In brief, even if few comparisons can be established between the present and previous research, a common finding that this study supports is that vocabulary acquisition can take place through exposure to multimodal input, and that this holds true across different proficiency levels. Taking into account the differences in design with previous studies, the amount of gains reported here seems reasonable, in spite of being still low. Lastly, it is also important to remark that the gains observed at all levels may have been underestimated, since learning of untested vocabulary may have taken place. Additionally, in the three proficiency groups, TV viewing may have also helped in enhancing other types of lexical knowledge (e.g., pronunciation of word forms, collocations, etc.) or contributed to develop partial knowledge of some words, which the tests did not capture.

8.1.6. Conclusion.

To sum up, this thesis has shown that vocabulary acquisition from TV series, combined with some explicit instruction, is possible at different proficiency levels and can be explained on the basis of the tenets of the DCT and the CTML. Besides, the CLT seems to be especially relevant in low-level learners. It is very important to bear in mind the decisive role that FL proficiency plays in this learning process. While intermediate learners seem to almost always benefit from exposure to multimodal input, beginner learners only appear to benefit from such exposure when they have accumulated more hours of extensive viewing; this may be probably due to their more limited TL proficiency and their constant need for extra help. In contrast, viewing subtitled TV series does not seem to aid upper-intermediate learners, possibly because of their already acquired TL proficiency, which allowed them to learn the target vocabulary from the classroom activities devoted to familiarising them with the TWs (as the CG did as well). Moreover, it is important to highlight the relevance of sustained exposure to multimodal input, as more significant differences between conditions arose towards the end of the pedagogical intervention, in part explained by the accumulation of input and the task familiarity effect. Finally, the combination of viewing TV series and formal regular instruction used in this dissertation led to similar or larger vocabulary gains as compared to previous research on the topic, even if differences in design between the present and other studies need to be accounted for.

8.2. Long-term retention of vocabulary learned from subtitled TV series

Data from Grade 6 and Grade 10 threw light on long-term vocabulary retention and showed that having been exposed to multimodal input for an academic year did not have a significant impact on the retention of the vocabulary learned during the last term of the intervention. In this respect, it seems that the extra help of TV viewing was not enough to retain, for several months, the knowledge acquired. However, the fact that both the EGs and CGs still remembered around 50% of the knowledge learned during the intervention, appears to indicate that, in any case, EGs' results were not worse than those obtained by students not watching the TV series.

First of all, the inclusion of a delayed test meets the call for further research on vocabulary decay (Barclay, 2018) and also takes into account some of the limitations that previous research hinted at (Frumuselu, 2015). As pointed out in the Literature Review (see Chapter 4), very few studies have investigated vocabulary retention from multimodal input. To our knowledge, only on four occasions (Baltova, 1999; Lertola, 2012; Nagira, 2011; Peters & Webb, 2018) have researchers investigated retention of vocabulary learned after video viewing. However, only in the most recent study, authors compared video viewing to no video viewing. In all the others, participants viewed videos under different subtitling conditions (traditional, bimodal, reversed subtitling, etc.) and the authors aimed at seeing which condition led to the highest degree of retention.

Another issue that needs some consideration is the time period between immediate and delayed testing. Vocabulary research has suggested that delayed post-tests should be taken at least two or three days after the immediate post-test (Nation & Webb, 2011). Besides, others label as ‘long-term retention’ those tests administered two weeks after immediate testing, although “the same period of time is but an instant in the natural vocabulary development process” (Gu, 2003, p. 12). In multimodal input studies, authors opted for including a delayed post-test administered between one and two weeks after the regular post-test. In our case, the test was administered eight months after the intervention, making a very large difference in the time span. Hence, we do need to distinguish between short-term retention (what previous studies investigated) and long-term retention (what this dissertation aims at measuring).

The index calculated from the delayed post-test considerably differs from previous research too. While past studies basically compared the results of the immediate and delayed post-test, the present study has only taken into consideration the target vocabulary which was actually learned during the intervention, excluding what was already known prior to the beginning of the term. We have later classified this vocabulary as maintained or lost eight months afterwards.

8.2.1. Multimedia and cognitive load theories: L2 vocabulary retention.

Results of this doctoral dissertation do not contradict the DCT (Paivio, 1986) and the CTML (Mayer, 2009) in Grade 10 and in the last term of Grade 6. As pointed out, both theories claim that language will be better learned if different modes of input are combined. It is also reasonable to think that, if the two theories are based on the idea that multimedia exposure contributes to greater depth of processing and better recall, knowledge would be more easily retrieved after some time too. However, if no exposure of such type is provided during a long period of time (in this case, eight months), the benefits will be lessened and may even disappear. The central idea of the DCT, which builds on the referential connections between logogens (as verbal representations) and imagens (as visual representations), seems to lose momentum in delayed testing since only logogens, and not imagens, could have been activated, as no visual input related to the TWs was received prior to the completion of the delayed post-test or the test itself. The same can be said about the CTML: as has been said throughout the thesis, its central idea is that people learn better from verbal and visual information than from verbal information. Since its basic tenet cannot be accomplished between immediate and delayed post-testing, it is reasonable to believe that the benefits it presupposes will not be obtained.

The results from the delayed test showed that the vocabulary that was learned from exposure to multimodal input was not better retained than that learned thanks to explicit teaching. This does not falsify the theories which frame the study as we have seen they are applicable to immediate vocabulary learning, but it limits their extent. We could claim that the theories that once supported the learning of vocabulary from multimodal input do not seem to hold to explain vocabulary retention in the long-run. This could be linked to the fact that, although it was not possible to know what the participants did between the end of the intervention and the administration of the delayed test, the chances of encountering the same TWs in other TV series or in movies are not very high. If they had been exposed again to these TWs, it is possible that such exposure would have been through a written source of input (e.g., textbooks, books, websites, etc.), and not through multimodal input, although most probably participants did not encounter many of the items again. Possibly, if the vocabulary gains made by the EGs had been larger or there had been more exposure to TV series (e.g., with episodes lasting around 40 minutes each), better retention rates could have been observed. Nevertheless, in the conditions of the study (more than eight hours of video viewing) and with the gains found, such conclusion cannot be drawn. Undoubtedly, more research would be needed on the issue of whether multimodal input leads to better vocabulary retention.

8.2.2. Factors affecting L2 vocabulary retention.

When scrutinizing the results on retention, we notice two facts: on the one hand, the percentage of knowledge retained is similar in both experimental conditions. To be precise, in Grade 10, the differences between the vocabulary remembered by the two conditions were 6.97% (form) and 13.63% (meaning), while in Grade 6 these were 1.45% (form) and 0.61% (meaning). At first sight, one may think that formal classroom instruction alone led to similar degrees of retention than video viewing at the two levels. In addition, there were no statistically significant differences between the two conditions, with the significance values ranging from .217 to .455 in Grade 10 and from .754 to .892 in Grade 6. This then confirms that none of the approaches is more beneficial for EFL learners than the other.

On the other hand, it is also true that those who lose more are those who had obtained more gains (i.e., the EG). Due to the higher scores the EGs obtained in Term 3 (relative gains between conditions proved to be statistically significant at T3 and in both proficiency groups), they had to retain a larger number of forms and meanings while the

CG had to remember fewer, sometimes not even half. As our memory capacity is limited (Baddeley, 1992), it is inherently more difficult to remember ten items than to remember three, so, in terms of mental effort and task demands, the intellectual exercise that the EGs had to perform was more challenging than that of the CGs.

Therefore, it should be noted that the percentage of vocabulary remembered may be misleading as it did not take into account the number of TWs that learners were asked to retain. Another formula would need to be applied or different analyses should be conducted to control for the number of items previously learned. In this respect, it would make sense to use a series of ANCOVAs with the number of TWs that needed to be remembered as covariate. Nevertheless, these cannot be run because “the independence of the covariate and the treatment effect” assumption is not met (Field, 2009, p. 397). That is, in order to use ANCOVA, the covariate should be completely independent from the treatment effect because “when treatment groups differ on the covariate then putting the covariate into the analysis will not ‘control for’ or ‘balance out’ those differences” (p. 398). This assumption can be easily checked with a *t*-test or an ANOVA: “if the groups do not significantly differ, then we can use the desired variable as a covariate” (Field, 2009, p. 399). If these confirmatory analyses are run, significant differences arise between conditions on the number of word forms and word meanings which needed to be remembered in the two proficiency groups, further proving that one-way ANCOVAs were not appropriate in this situation (Miller & Chapman, 2001). Besides, the potential covariate was not measured prior to the beginning of the intervention and the scores on the two testing times are strong and significantly correlated, confirming that the covariance analysis could not be conducted (Pallant, 2013).

The percentage of forms and meanings remembered could also be influenced by the uncontrolled and random exposure to the TWs learners may have had in the eight months between post-test T3 and delayed test. It is true that some of the TWs were infrequent and hard to encounter in non-instructed contexts (e.g., *to flinch*, *ply*, *to schmooze*, *zit*...), others, though, could be found, for instance, in the following year's coursebook or in out-of-class activities (e.g., videogames, books and magazines, etc.). Besides, it cannot be ruled out that some of the participants kept on watching the same TV series after the end of the intervention and encountered some of the TWs again, it being the most optimal context for new encounters. Because research has shown that the more encounters with a word, the higher the chances to acquire it (Brown et al., 2008; Pellicer-Sánchez & Schmitt, 2010; Rott, 1999; Uchihara et al., 2019, among others), this extra exposure provided by other activities could have been conducive to more learning and could have been particularly beneficial for those TWs that were partially learned during the intervention.

In connection with this, the role that vocabulary recycling plays in lexical development needs to be considered as well. As shown by research, "new words need to be recycled regularly to be learnt" (Schmitt & Schmitt, 1995, p. 139), with many revisits to the target items and an increasing spacing between them (Nation, 2013). Vocabulary recycling was very limited throughout the intervention, as the TWs were not usually encountered in the episodes in which they were not target. Actually, in Grade 6, the TWs in Term 3 appeared 7.40 times on average in the episodes in which they were target, and 9.40 times if all the episodes of the term are considered. Thus, TWs only appeared twice in the other seven episodes of the term. As far as *Seinfeld* is concerned, TWs appeared 6.37 times on average in the episode they were considered target vocabulary and 6.60 times when the seven episodes of the term were inspected, meaning that the TWs only appeared 0.23 times in the remaining six episodes. Hence, vocabulary was not often recycled, and the authentic materials selected for the intervention in Term 3 did not give students the chance to regularly come across the TWs in other episodes. In addition, as explicit attention was not directed towards the items when they were not considered to be TWs, chances are that learners did not notice them. Possibly, if vocabulary had been consciously recycled during the months after the intervention, retention rates could have been higher.

As pointed out by Peters and Webb (2018), retention can also be attributable to higher levels of proficiency at the time the delayed test was administered as compared to when post-test T3 was taken. Particularly in Grade 6, where there still was much room for improvement, learners might have developed more metalinguistic awareness that facilitated the acquisition of new vocabulary (Bae, 2015).

It is also important to remark that the control condition, which was not exposed to subtitled TV series and learned the TWs from deliberate learning, also retained more than half of the TW forms and around 50% of the TW meanings. As participants knew that they would be tested on the items they encountered in the vocabulary pre-task, it is possible that they interacted with the TWs more actively during the pre-teaching phase (they knew they did not have many more opportunities to learn the target vocabulary). For instance, they could have related the items with personal experiences, taken notes of what the teacher or researcher was explaining, or made an extra effort to deliberately learn the words. This more active and meaningful interaction (for them) could have contributed to transfer knowledge from participants' WM to their LTM, thus facilitating vocabulary retention (Schmitt, 2000).

Finally, the role of feedback may also account for the vocabulary retained by the CG in the delayed test. Feedback, or the "information given to learners regarding their performance" (Nakata, 2014, p. 416), is said to facilitate FL learning and its effects are maintained over time (Li, 2010). In this sense, learners in the CG (as well as those in the EG) made their hypotheses about the forms and meanings of the target items when they first encountered them and tested such hypotheses in the pre-task. Immediately afterwards, they received feedback on their performance (pre-tasks were corrected in class), which no doubt contributed to vocabulary learning. Hence, even if we have seen that exposure to multimodal input led to vocabulary retention, so did deliberate teaching, possibly due to a more active processing of the TWs in the pre-task and careful attention to the feedback learners received.

8.2.3. Interpreting retention rates in relation to previous research.

As research on vocabulary learning from TV viewing is quite recent, its long-term effects regarding vocabulary retention have not been studied much in depth. Again, comparisons with previous research are difficult to make. Some of the scarce research on the topic found traces of learning between immediate and delayed testing. In addition, some studies also showed that subtitles did not contribute to higher retention, in line with what was observed in the present thesis (e.g., Lertola, 2012; Nagira, 2011). However, none of the studies mentioned below exclusively focused on the knowledge acquired during the experiment (words already known were also considered) and did not calculate the percentage of knowledge gained and retained, which makes comparisons a bit difficult to establish.

Baltova (1999), for instance, found that the students watching subtitled videos (reversed and bimodal) retained less vocabulary than the CGs, which goes in line with the fact that the CG in Grade 6 retained more word forms and meanings than the EG. The retention rates found in her research were much higher than those of the present dissertation: the subtitled conditions remembered 84% of what they knew on the post-test and the traditional condition learned some vocabulary[10]. This higher retention rates are possibly attributable to the fact that there were two weeks between immediate and delayed testing and that the number of items to remember was also smaller (30 vs. 40 words). The results of the thesis go also in line with Nagira (2011), who showed that there were no significant differences between the captioned and the uncaptioned conditions. The retention rates found in the present dissertation are also much lower than those found in her study (98% –captioned condition– and 99% –uncaptioned condition–); nevertheless, her students took the delayed test only one week after the end of the treatment and were only tested on 12 TWs in a VKS format. Lertola (2012) also found deliberate learning between immediate and delayed testing in both experimental conditions. The EG, which created the English subtitles for a given video, remembered a 107% of what they knew in the post-test, and the CG, which did other tasks unrelated to the video, remembered a 101%. As can be seen, these retention rates are not comparable to ours, but we need to bear in mind that both the amount of target vocabulary (15 vs. 40 TWs) and the test format (VKS vs. form and meaning recall) also differ a great deal. Finally, Peters and Webb (2018) did not analyse the delayed tests (meaning recognition and meaning recall) due to the high indices of deliberate learning observed at this testing time, so retention rates cannot be compared. The authors showed that more than half of the correct responses in the delayed tests were

[10] The percentages presented in this and the following paragraphs have been recalculated by the researcher so as to be able to compare retention rates across different studies.

newly learned items, which they attributed to deliberate learning, and to test and guessing effects[11].

Although their study did not involve video viewing, Yoshii and Flaitz (2002) explored vocabulary retention through different input modes. They compared the vocabulary learned across three annotation types (text-only, picture-only, and a combination of the two) among beginner and intermediate EFL learners. Participants were asked to read a story, specially written for the purpose of the study, containing 14 TWs for which they could consult either a written definition, a pictorial representation or both. Students were asked to complete a definition supply, a picture recognition and a MC meaning recognition test at two points in time: immediately after reading and two weeks later. All the tests were unannounced. The authors claim that, two weeks after the end of the experiment, significant differences in favour of the combination group were only found for the picture recognition test. What is surprising is the high indices of retention observed on the three tests by the group exposed to written and visual input: 86.86% on the picture recognition, 79.95% on the word recognition and 61.29% on the definition supply. However, the first two tests were much easier than that used in the present study. Furthermore, and bearing in mind that only two weeks passed between the two testing times, the retention rates on the definition supply test can be said to be quite similar to ours.

[11] Even though there are more studies on video viewing which administered a vocabulary post-test one or more weeks after the end of the experimental phase (e.g., Aurstad, 2013; Kvitnes, 2013; Zarei, 2009), they have not been referred to in this section since they did not additionally include a post-test just after viewing the videos. This makes comparisons impossible to draw because we make a difference between (a) what students learn throughout the pedagogical intervention and (b) what they retain eight months after the treatment ends, whereas these other studies only look at what students learn since the beginning of the experiment.

Research on reading has also evaluated vocabulary retention and it is interesting to check retention rates from unimodal input against those from multimodal input. Two well-known studies are Rott (1999) and Waring and Takaki (2003). The former explored the acquisition and retention of words encountered different times by intermediate university learners. The first delayed post-test was administered one week after the end of the treatment and the second, four weeks later. All of them were meaning supply tests in which learners were instructed to provide the meanings of all target items[12]. After proving that more exposure to the TWs led to higher levels of acquisition, the author showed that more learning also implied less retention, especially on the second delayed test. In relation to the first delayed test, retention rates showed that there was no loss of the vocabulary previously acquired through reading. The retention levels in the study (31% on average in the first delayed test and 24% in the second) are much lower than the percentage of meanings remembered by our Grade 6 and Grade 10 students, probably because of the nature of Rott's study and the fact that a different modality of input was involved (reading vs. video viewing). In another piece of research, Waring and Takaki (2003) analysed the acquisition and retention of a set of invented TWs after reading graded readers. University participants were asked to complete word-form recognition, meaning translation (similar to our pre- and post-tests) and meaning recognition tests (similar to our post-task) immediately after reading, one week later and three months later. Results showed that the participants forgot many of the words they had learned. To be precise, three months after the end of the experiment, they remembered the 54.90% of the words learned on the form-recognition test (considered to be the easiest one), 57.55% on the meaning recognition test and 19.57% on the translation test. They also showed that retention was also influenced by the number of exposures to the TWs: the more exposures learners had had

[12] Other receptive vocabulary tests were administered (e.g., a select-definition task), but they will not be referred to in the discussion since no comparisons could be drawn.

during the reading phase, the higher the chances to retain this vocabulary. Focusing on the meaning translation test (the closest resemblance to our delayed test), we see that our participants retained much more knowledge (48.81% in Grade 10 and 56.86% in Grade 6) than in Waring and Takaki (2003), even if they were at lower proficiency levels and the timespan was also longer (three vs. eight months). Nevertheless, it must be taken into account that both Rott (1999) and Waring and Takaki (2003) did not include any explicit teaching of the target vocabulary, which made the learning process truly incidental. What is more, in the second study, the TWs were invented, yet plausible, items, so no vocabulary recycling could occur during the three months between immediate and delayed testing. Likewise, the nature of the two modalities of input are radically different (TV series vs. short written paragraphs or graded readers).

8.2.4. Conclusion.

In sum, results show that being exposed to multimodal input does not necessarily imply better word retention and that those who learned more during the experiment were not fully able to remember that knowledge. When analysing the two conditions separately, the higher learning rates experienced by the EGs in Term 3 influenced the scores on the delayed test: since learners in the EGs had to remember much more than the CGs, the task was a bit more complex for them, in comparison to the other experimental condition. This could be a possible explanation for the non-significant differences between the two groups. Finally, retention rates examined in this dissertation are not comparable to those observed in previous research on vocabulary retention from multimodal input, which mainly tapped into short-term retention. Overall, it appears that short-term retention is higher and has been more researched than long-term retention, and that video viewing is conducive to similar (e.g., compared to visual and written annotations) or higher (e.g., compared to reading) retention rates than other modalities of input.

8.3. Content comprehension of subtitled TV series

Results on content comprehension showed that it did not improve throughout the academic year. Besides, high variability within a single term was also observed and the scores on the comprehension tests varied a lot depending on each episode and participant. Inferential tests showed a significant main effect for time in most cases, but this was probably due to the above-mentioned variability. It was seen that the mean comprehension score obtained by university participants was significantly higher than that of sixth graders, but not higher than that of tenth graders. In contrast, tenth graders significantly and constantly outperformed sixth graders. These results at each level mirror those obtained for lexical gains.

8.3.1. Lack of improvement in content comprehension.

The fact that there is no clear and positive evolution in the comprehension scores throughout the intervention in any of the three proficiency groups could be somehow expected. In the next sections, we will point out different reasons and argue why they may or may not explain this lack of improvement in learners' content comprehension: these are, namely, limited exposure to audiovisual input, test type and test difficulty, vocabulary demands of the episodes and some other factors that we could not control for (e.g., reading speed or participants' interest in the episodes).

8.3.1.1. Amount of exposure to multimodal input.

Learners from the EGs in Grades 6 and 10 were exposed to a bit more than eight hours of multimodal input in one academic year. Although this is quite a lot of extra input in our curricular system, eight hours are not a considerable amount of exposure in comparison to the input people are exposed to in daily life. For instance, eight hours of authentic input is the same that one can be exposed to by living with a host family in the UK or the USA in one single day. Consequently, not enough input was provided for content comprehension to significantly improve. In relation to this, L2 listening comprehension has been said to be the least explicit of the language skills and, probably, the most difficult to learn (Vandergrift, 2004), so more time would have been needed to see some progress in participants' content comprehension.

The reasons that explain the poor, yet significant, lexical gains are also true for content comprehension. Another way to assess whether extended exposure to TV series increased content comprehension could have been by choosing two episodes with exact coverage levels to be seen in September and June and then compare the results obtained. However, two episodes with exactly the same difficulty that engage all learners in exactly the same way are difficult to find in practical terms, and even in this case, we cannot guarantee that the two are equally demanding.

8.3.1.2. Test type and difficulty.

One may think as well that this lack of significant improvement is due to the tests we used and / or their difficulty. However, there are several reasons that rule out this possibility.

First of all, tests were carefully aimed at measuring content comprehension, instead of L2 reading skills, so that is why they were designed and administered in one of the participants' mother tongues (Spanish), as opposed to the TL. Second, the comprehension tests included three types of exercises which were different and took into account the distinct facets of listening comprehension (Buck, 2001), resulting in a very comprehensive test: from more detailed questions tapping into quantities, characters or specific information to a final exercise that evaluated global comprehension of the episode.

Furthermore, there was no relationship between the difficulty indices of the comprehension tests and the scores learners obtained in them (i.e., 'easy tests' did not result in higher comprehension scores), proving that test difficulty cannot explain the results and the episode variability. If we closely analysed one of the terms in which pairwise comparisons showed significant differences between the first and the last episode –T3 in Grade 6– (where the final episode was significantly higher than the initial one), we will see that this evolution cannot be solely attributed to the difficulty of the tests. Actually, in that term, both initial (E17) and final (E24) tests were of 'moderate' difficulty and participants scored significantly higher in E24 than in E17.

In addition, test difficulty was shown to be appropriate across proficiency groups: instances of 'easy' and 'moderate' tests were found in all three groups, and none of them proved to be 'difficult' or 'very difficult'. Furthermore, there was no ceiling effect in any of the tests. The highest score obtained by university learners was 88.65% in E03, which means that, on average, participants could still get two more questions right. The same holds true in Grade 10, since the highest score was 82.76% (E02), and in Grade 6, where the maximum score was 65.76% (E09). At all proficiency levels, learners could score higher and there was room for improvement. This also evidences that tests were at the appropriate level and were very reliable and that the lack of improvement in comprehension cannot be attributed to the tests designed and administered.

8.3.1.3. Vocabulary demands of the episodes.

Another possible explanation for not finding statistically significant changes in content comprehension along the intervention could be that comprehension scores were higher or lower depending on the difficulty of the vocabulary in each episode. The close link between comprehension and lexical knowledge (Laufer & Ravenhorst-Kalovski, 2010; Pulido, 2004; Qian, 2002; van Zeeland & Schmitt, 2013b) might lead us to think that more lexically demanding episodes would entail lower scores. However, that was not the case in the present study: there were no differences in the lexical coverage of the episodes and lexical profiles were comparable, as will be shown with examples from the TV series.

To start with, 95% lexical coverage was reached at the 2k level in all the eight episodes to which university participants were exposed. In addition, at the university level, there were significant differences between the scores on E03 and two more episodes (E06 and E07), but these were not easier or more difficult than the rest in terms of vocabulary demands (i.e., 98% coverage was reached at the 3k level in E03 and E07 and at the 4k level in E06). Continuing with *I Love Lucy* (T2) and *Seinfeld* (T3), the lack of a relationship between lexical coverage and comprehension scores was evident as well. The mastery of the 2k frequency band granted participants 95% coverage in fifteen episodes, while in the remaining seven (E11, E15, E17, E18, E20, E21, E22), three thousand words were needed. Again, these do not correspond to the most difficult seven episodes. To serve as an example, Grade 10 participants obtained the third highest score of the year in E22. Moving on to *The Suite Life of Zack and Cody* and *Wizards of Waverly Place*, knowledge of the first two thousand words of the English language sufficed in 13 episodes, whereas three thousand words were needed in nine more and, in only two episodes (E08 and E23), four thousand words had to be mastered to understand 95% of what was being said. Despite this fact, these two episodes were not specially challenging for participants; actually, E08 obtained the third highest score in Term 1 and E23 did not correspond to the lowest score in Term 3 either.

Regarding the lexical coverage of each of the quarterly corpora built with the transcripts from the different episodes, in Grade 6, 95% coverage was reached at the 3k level in Terms 1 and 2 and at the 2k in Term 3. If we take the 3k level as the reference point, we will see that the coverage levels were almost identical: T1 → 95.89%, T2 → 95.83%, T3 → 96.28%. In Grade 10 and university, knowledge of the first two thousand words of the TL sufficed in Terms 1 and 2 and mastery of the 3k band was needed in Term 3. Taking the former as the reference point, we do find that coverage levels at the 2k level are quite similar as well: T1 → 96.85%, T2 → 95.72%, T3 → 94.95%.

In the following graphs, what has just been explained is visually represented, and it is made clear that the lexical profiles of the episodes are similar, as the different lines (one per episode) are hardly distinguishable from one another. To facilitate understanding of the graphs, frequency bands have been grouped into high-frequency words (1k and 2k), mid-frequency words (3k to 9k) and low-frequency words (10k to 25k), following Nation (2013). Off-list words (e.g., *schoolboy*) have not been included in the low-frequency group since it cannot be taken for granted that they are as infrequent in the English language as other words which have been indeed classified between the 10k and the 25k bands (e.g., *bellhop* or *dime*).

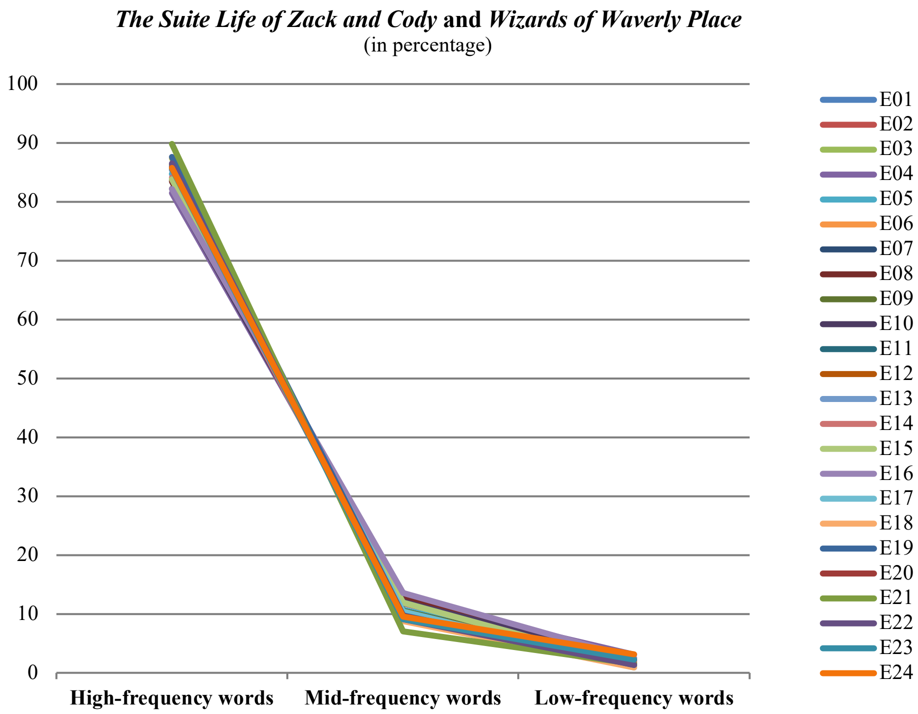


Figure 8.1 - Lexical profiles of the episodes of *The Suite Life of Zack and Cody* and *Wizards of Waverly Place*, divided into high-, mid- and low-frequency words.

Figure 8.2 - Lexical profiles of the episodes of *I Love Lucy* and *Seinfeld*, divided into high-, mid- and low-frequency words.

We have seen that the vocabulary demands of the episodes used in the intervention did not differ much between them. It is also important to compare them with previous research on lexical coverage of television programmes (Webb & Rodgers, 2009b) so as to see whether they were a good representation of the input people are exposed to when watching English language television. Webb and Rodgers selected both British and American television productions from different genres. However, for comparison purposes, only American programmes will be considered here, as the four TV series used in the study were American productions and were originally aired in the United States. Regarding the television genres, children's programmes were selected as the reference point for Grade 6 learners. Nevertheless, two different genres were chosen for Grade 10 and university participants: older programmes for *I Love Lucy* and sitcoms for *Seinfeld*, precisely because two episodes from each TV series were analysed by the authors (although those analysed in the article were not used in the pedagogical intervention). It should be mentioned as well that the authors examined the transcripts with the RANGE software (Nation & Heatley, 2002), whereas our transcripts were analysed with VocabProfile (Cobb, ongoing). Hence, the transcripts in Webb and Rodgers (2009b) were broken down into 14 frequency lists, plus proper nouns and marginal words, while ours were divided into 25 frequency bands, plus off-list words. In our analysis, proper nouns were categorised as 1k words (not in a separate frequency band) and marginal words were deleted from the transcripts. In the paragraphs that follow, knowledge of proper nouns and marginal words is already included in the coverage figures reported (as the former were classified in the 1k band and the latter were not taken into consideration).

Both Webb and Rodgers (2009b) and this doctoral dissertation show similar lexical demands of the television programmes analysed, with some minor differences depending on the television genre. In Webb and Rodgers (2009b), learners could have attained 95% coverage with knowledge of the first three thousand words (95.31%), including marginal words and proper nouns. This matches the results of the present study, in which knowledge of the first three thousand words was enough to reach 95% coverage in three of the corpora and, in the other three, the 2k band was enough. More differences were found regarding 98% coverage, as Webb and Rodgers showed that knowledge of the first seven thousand words was needed, whereas in our study, knowledge of the first five frequency bands was enough in two of the terms, always bearing in mind that knowledge of marginal words and proper nouns is included in such coverage figures.

Focusing on the different genres, as seen in Table 8.1, *The Suite Life of Zack and Cody* and *Wizards of Waverly Place* turned out to be slightly more demanding than Webb and Rodgers' children's programmes. In contrast, *I Love Lucy* was shown to be slightly easier than the older programmes they analysed. Finally, no differences were observed between *Seinfeld* and other sitcoms.

Table 8.1
Frequency bands at which 95% and 98% lexical coverage were reached in Webb and Rodgers (2009b) and in the present study.

	Coverage level	Webb & Rodgers (2009b)	PhD thesis
Children's programmes (*The Suite Life of Zack and Cody* and *Wizards of Waverly Place*)	95%	2k	3k
	98%	5k	6k
Older programmes (*I Love Lucy*)	95%	3k	2k
	98%	6k	5k
Sitcoms (*Seinfeld*)	95%	3k	3k
	98%	6k	6k

In conclusion, this proves that the episodes that were selected for the experiment were a good example of what people are exposed to when watching American television. In relation to the comprehension scores, we have been able to prove that the episodes selected for the intervention were not easier or more demanding than other programmes from the same television genres. Furthermore, we can claim that these episodes were adequate (they were not very difficult or very easy) for learners' proficiency level in the TL. Hence, variability in comprehension does not seem to be due to the vocabulary demands of the episodes, as regularity was observed between all the episodes selected for the intervention, and also in comparison to other television programmes with similar characteristics.

8.3.1.4. Other uncontrolled factors.

In the previous sections, we have argued that tests or vocabulary demands of the episodes cannot account for the lack of improvement in comprehension, but the lack of a clear pattern in the scores can otherwise be explained by a series of aspects that were not controlled for in the pedagogical intervention. To name a few, it is possible that some of the episodes may have been more interesting than others and that students were more attentive to them since, for some reason, they were catchier and less boring. Consequently, these more interesting storylines could have led to higher comprehension scores, as more attention would have been devoted to them. Additionally, speaking rate and accentedness could also make a difference. Some episodes could have been less intelligible than others because characters spoke faster or with an accent participants did not fully understand. For instance, in *The Suite Life of Zack and Cody*, two of the secondary characters were a German hotel director and a Latin American bellhop or, in *I*

Love Lucy, Lucy's husband (one of the main characters in the show) is from Cuban origin. Unfortunately, participants were not explicitly asked to rate their self-perceived degree of difficulty or their interest in each of the episodes, so these could not be analysed in relation to comprehension scores.

8.3.2. The role of proficiency in the content comprehension of TV series.

The results from the third RQ clearly show that the higher the proficiency level, the better the content comprehension of TV series. In relation to previous research linking proficiency and video comprehension, many studies included participants from different proficiency levels (e.g., Hayati & Mohmedi, 2011; Lavaur & Bairstow, 2011), but not many have been able to compare the results from two proficiency groups. Among those that have, Winke et al. (2010) found that the main effect for proficiency was statistically significant (p = .00) when analysing the content comprehension of second- and fourth-year university learners of Russian and Spanish. Similarly, Park (2004) found that "the more advanced the participants' English proficiency is, the better they are in handling both the aural and captioned information" (p. 113). To claim so, he draws on the fact that his above-intermediate group significantly outdid the pure- and below-intermediate counterparts (p = .000), and the middle group also performed significantly better than the least proficient group (p = .013). Eventually, the studies by Montero Perez et al. (2013) and Montero Perez et al. (2014) revealed a positive relationship between participants' VS and comprehension tests' scores, indicating that "the larger one's vocabulary, the better the comprehension score" (2014, p. 129). This is further supported by the fact that VS is closely related to general proficiency level and to the four language skills (reading, listening, writing and speaking) (Miralpeix & Muñoz, 2018).

Metacognitive abilities can also play a role in the higher comprehension scores achieved by more advanced learners, as opposed to less proficient students. University and secondary school students had more experience in L2 reading and listening tasks and were more skilful than Grade 6 students in paying attention to the audio and the subtitles at the same time. Sixth graders may have had difficulties in reading the subtitles (even if they were in one of their L1s). These typically require reading speeds that "hover around the 180 Words Per Minute (WPM) or 15 to 17 characters per second" (Díaz Cintas, 2010, p. 345) and beginner learners who participated in the intervention could read at an average rate of 154 WPM in Spanish, which falls short to be able to read the subtitles in full. Tenth graders, however, had a reading speed of 182 WPM in the L2[13], which facilitated reading of the subtitles and making sense of what was being said. Nevertheless, it is true that we do not know the extent to which participants actually read the subtitles, as eye-tracking data was not collected and participants were not interviewed[14]. It is possible that, towards the end of the year, participants relied less on L1 / L2 subtitles, but this did not lead to better comprehension (or, at least, we have not been able to capture it with the test that was administered). We could have shown them an episode without textual support from time to time to see their level of comprehension when subtitles were not available; however, this was not possible since all participants were provided with L1 (primary) or L2 (Grade 10 and university) subtitles from the beginning to the end of the intervention.

These differences in proficiency may explain why high-school students constantly and significantly outperformed sixth graders, even if Grade 10 series were subtitled in the L2.

[13] Both scores (L1 and L2 reading speed) were taken from different tests administered in the SUBTiLL Project. The L1 reading fluency test was adapted from Galí (1984) while the L2 reading fluency test was adapted from Quinn, Nation, and Millett (2007).
[14] See Winke, Gass, and Sydorenko (2013) for an example of research on caption-reading behaviour by FL learners.

Eventually, university and high-school participants' content comprehension did not statistically differ: although they were different in terms of general proficiency (university learners knew 4,431 words while tenth graders' VS was of 3,498 words), their L2 reading speed, their metacognitive ability to understand the input and their development of comprehension strategies are probably very similar. This can also be applied to their learning autonomy and self-regulation.

8.3.3. Interpreting degree of content comprehension in relation to previous research.

In this section, the content comprehension scores found in this thesis will be interpreted in the light of previous research on the topic. To do so, previous studies' results were recalculated by the researcher based on the mean and maximum scores reported by the authors. For comparison purposes, only the scores from those groups that were exposed to multimodal input under similar conditions as our EGs will be referred to. That is, the scores of those learners exposed to L2 audio and L2 subtitles will be compared to results found in Grade 10 and university. In contrast, the scores of those participants exposed to L2 audio and L1 subtitles will be outlined when comparing these to Grade 6 results. The scores from other groups in different conditions (e.g., L1 audio and L2 subtitles) will not be reported.

Regarding upper-intermediate learners, only one study will be reported here (Garza, 1991) since the other pieces of research with this population that have been identified (Brett, 1997; Lavaur & Bairstow, 2011; Winke et al., 2010) were very different from the present study and were not comparable. Garza (1991) found that low-advanced university students obtained a mean score of 87.10%, after answering 10 MC questions, with three options each, about five distinct video segments. This is not that different from the mean comprehension score seen in the dissertation (81.47%), although the test formats are not the same (MC questions as opposed to T/F, MC and ordering statements) and Garza's study was based on videos lasting no more than 18 minutes.

As can be seen in Table 8.2, comparisons with previous research on intermediate learners and content comprehension from audiovisual materials are quite difficult to establish. First and foremost, most of the studies summarised in the Literature Review were one-off research pieces, and they included a variety of questions (T/F, MC, matching, open-ended) and comprehension tests. What is more, most studies only included one comprehension test, and not a battery of them, administered just after the videos were seen. Bearing these methodological differences in mind and treating all comparisons with caution, most studies identified in the thesis have found lower degrees of content comprehension than those achieved by Grade 10 learners. Nevertheless, if only Rodgers (2013) is taken into account (as his study was similar to ours), the mean comprehension score is quite close to the results we obtained (66.51% vs. 69.34%), shedding more light on the degree of content comprehension that intermediate learners may achieve with this type of input.

Table 8.2
Results of the studies on content comprehension with intermediate learners compared to the results of the doctoral thesis.

Study	Type of test	Input	Comprehension results (in percentage)
Baltova (1999)	Open-ended questions	Short video (7m 33s)	39.18% on the post-test 27.64% on the delayed post-test
Markham et al. (2001)	MC questions	7-minute DVD	56.70%
Park (2004)	Gap-filling activities, MC questions, word recognition test	Six short clips (2-4m each)	61.37%
Hayati & Mohmedi (2011)	MC questions	Six short clips (34m in total)	66.37%
Rodgers (2013)	T/F and MC questions	10 episodes of a TV series (+7h)	66.51%
Rodgers & Webb (2017)	T/F and MC questions	10 episodes of a TV series (+7h)	66.51%
Montero Perez et al. (2013)	Global, detailed, open, closed, receptive and productive questions	Three short clips (16m in total)	69.33%
PhD thesis	**T/F, MC questions and ordering statements**	**22 episodes of a TV series (+8h)**	**69.34%**
Montero Perez et al. (2014)	T/F, open-ended and matching questions	Three short clips (10m in total)	71.49%
Huang & Eskey (1999)	MC questions	One episode of a TV series (21m)	72.50%
Kvitnes (2013)	MC test matching unfinished sentences	One episode of a TV series (20m)	89.11%
Aurstad (2013)	MC test matching unfinished sentences	One episode of a TV series (20m)	93.40%

Generally speaking, the degree of content comprehension at Grade 6 seems to be a bit below what other studies had found, yet differences are not very large. The comprehension scores from previous research that will be listed are the result of a set of content comprehension questions based on a single episode, and usually administered immediately after the viewing. In the present dissertation, the score that has been taken as reference is the result of 24 comprehension tests, with three different exercises each, administered after the viewing of each episode. Hence, these comparisons should be taken with due caution, since it has been seen that there is a high episode variability and we do find instances of episodes with scores far above the levels observed in previous research (e.g., E07, E09 or E24). We should acknowledge that the interpretation of the present results may vary if there were other longitudinal studies on content comprehension available.

That said, there were only two studies with beginner learners from which percentage comprehension scores could be calculated. To start with, d'Ydewalle and Pavakanun (1995) conducted two experiments; in the first, beginner university students were exposed to multimodal input under nine conditions and were asked to answer 33 MC questions based on a 12-minute video fragment. Those that saw the audiovisual material with Dutch (L2) audio and English (L1) subtitles, managed to answer 72.22% of the questions right, far above the 52.73% comprehension obtained by Grade 6 participants. More similar comprehension scores to our study were seen in the second experiment, with Grade 9 learners. This group, who can be more readily compared to sixth graders, achieved 59.54% of content comprehension. However, the test was administered immediately after the viewing and it only included one type of questions (MC). It is important to remark that there was a three-year difference between this second group of participants and the beginner group that was included in the thesis, which could explain the 6.81% difference between the two cohorts.

Finally, Galimberti (2016) translated and slightly adapted one of the tests that were used in Grade 6 into Italian (the L1 of her participants) and found that Grade 6 learners of English exposed to English audio and Italian subtitles achieved 61.81% comprehension. This figure is nine points above the average comprehension rate observed in this thesis. Nevertheless, if the episode she selected (E05) is considered, the comprehension score achieved by the EG was 55.56%, which is closer to what Galimberti (2016) found.

8.3.4. Conclusion.

To sum up, it is possible that the lack of significant development of participants' content comprehension could be due to the limited amount of authentic multimodal input the EGs were exposed to, even if it was more than in their usual curriculum. In the same way that it is unreasonable to think that one's content comprehension will get better after one single day in contact with the English-speaking community, it also seems implausible that eight hours of extensive viewing will have a strong significant impact on EFL learners' content comprehension. Besides, the lack of a clear pattern in the comprehension scores may be due to the intrinsic nature of the episodes themselves. It has been proven that the lexical demands did not vary across the episodes and that test difficulty was similar throughout the academic year. Other intrinsic characteristics of the episode (e.g., storyline, self-perceived difficulty, degree of interest, intelligibility, etc.) may account for some variability in the comprehension scores, which is something common to find in research on the topic (e.g., Rodgers, 2013).

CHAPTER NINE - CONCLUSION

This PhD thesis aimed at investigating vocabulary learning through subtitled TV series, as well as its long-term retention, in three different proficiency groups (university students, secondary students and primary students). It also explored the extent to which these learners understood the TV series episodes and whether there were changes in comprehension after extended exposure to TV series. All learners (EGs and CGs) received explicit vocabulary instruction, but only half of them (EGs) were additionally exposed to L1 (primary) or L2 (high school and university) subtitled television episodes. To measure vocabulary learning, students were pre- and post-tested, at the beginning and the end of each term, on their aural and written recall of a set of TW forms and meanings. Besides, eight months after the end of the intervention, primary and secondary students were also administered a delayed test tapping into their knowledge of the TWs previously presented in Term 3. So as to investigate content comprehension, the EGs were asked to complete a comprehension test at the end of each viewing session.

Theoretically framed by the DCT (Paivio, 1986, 2007), the CTML (Mayer, 2002, 2009) and the CLT (Chandler & Sweller, 1991; Sweller, 1994), the dissertation wanted to see whether watching TV series led to better learning in combination with explicit teaching, as opposed to more traditional forms of instruction. Based on these theories and what previous research had found, it was believed that those learners who were additionally exposed to television series may do better on the vocabulary tests and may also retain more vocabulary in the long-run as new words would be processed more in-depth through different channels. Regarding content comprehension, it could be expected to increase along the academic year, as students get used to this practice and more input was being provided. However, and due to the lack of a CG, no comparisons could be established regarding vocabulary gains or content comprehension of the TV series without subtitles.

9.1. Review of findings

In light of the analyses run with the data collected for this doctoral dissertation, several findings should be highlighted. This thesis first showed that vocabulary learning takes place at all levels independently of the experimental condition, yet we need to bear in mind that vocabulary gains were quite small. Even if the intervention was one academic year long, the amount of input learners received was actually low in comparison with naturalistic contexts and it was not enough to find larger gains. More exposure and repeated encounters would probably be needed, as vocabulary learning is also a very slow and challenging process. Word forms were better learned than word meanings in all cases, although L2 subtitles did not particularly favour their learning, nor did L1 subtitles facilitate the learning of TWs' meanings in the low-proficiency group.

Differences between EGs and CGs seem to be mediated by learners' proficiency level. Intermediate level students at Grade 10 were the ones who benefitted the most from subtitled television programmes. At more advanced levels (e.g., upper-intermediate), adding video exposure did not make a difference in vocabulary learning. At beginner levels, students may be overwhelmed by the presence of too much information that hindered the learning and retention of new words, and they consequently fail in making the most out of multimodal input. However, it is also true that sustained exposure to FL television at all proficiency levels can bring about positive outcomes, as more gains were observed towards the end of the year. In the long-term (i.e., in the delayed test), participants from the two experimental conditions at the two proficiency levels tested eight months after remembered to the same extent the vocabulary learned during the intervention. This was probably because TWs were not encountered again, nor was this new vocabulary recycled in class.

Regarding content comprehension, no clear improvement was observed in any of the groups. Again, more sustained exposure to audiovisual materials would still be needed to enhance comprehension. Besides, the scores on the comprehension tests were participant- and episode-dependent, as it also happens in Rodgers (2013) or Pujadas (forthcoming). It seems that this was not caused by the vocabulary demands of the episodes or that the tests were not measuring comprehension accurately. Nevertheless, this variation in different episodes could be related to the participants' perception of the episode's difficulty and to the extent to which students found it more or less interesting and engaging.

Eventually, the important role that proficiency plays in language learning through multimodal input was ratified by the fact that university and secondary school participants significantly learned more vocabulary and obtained higher scores in content comprehension than primary school children, especially at the end of the intervention. Nevertheless, a three-year difference (between high-school and university participants) and some more previous exposure to the TL did not lead to more vocabulary learning and better comprehension of the TV series.

9.2. Pedagogical implications

These results have several pedagogical implications that are worth considering if the intervention wants to be implemented in other classroom contexts. First, to maximise vocabulary learning, it is enormously important that the video viewing experience is accompanied by a set of active learning tasks, which guide learners throughout the process. We have seen that, even if the pedagogical intervention made students focus on the target vocabulary with the vocabulary pre-tasks, the learning rates were somehow limited, especially at lower levels. Had not the vocabulary pre- and post-tasks been included, it is very likely that incidental vocabulary learning would not have taken place. Hence, it is highly advisable to include some warming-up tasks for each episode that include the TWs. Similarly, a consolidation task at the end of the session is advisable to revise the vocabulary and to provide learners with feedback. Without this set of focused tasks, incidental vocabulary acquisition at the proficiency levels analysed might just not take place, as no explicit attention would be allocated to the target vocabulary and the information would not be processed with the intention of remembering it. In addition, learners do not always notice unfamiliar vocabulary and, if noticed, chances are this would not be frequently reactivated (which, in the thesis, was achieved thanks to the vocabulary post-task, the TV series and the post-test) and forgotten sooner than later (Hulstijn, 2013). Overall, it is important to claim that unguided viewing is not recommended in our context, at least in a classroom setting at these ages and proficiency levels.

Second, we have seen that vocabulary learning through FL television, along with some focused instruction, can be positive for language learning. However, it has also been shown that comprehension does not improve along the academic year. Vocabulary acquisition seems to improve thanks to the presence of the video image, which provides learners with a visual representation of what is being said, and the textual support, which can help to parse the aural speech and to ease understanding. Hence, it would be recommendable to use subtitled FL television in the classroom context as a means to complementing and reinforcing more traditional vocabulary acquisition activities. To do so, though, all the agents involved (e.g., learners, teachers, learners' relatives) should be well informed about the potential that FL television has for L2 development (Webb, 2015). At first, there may be a general reluctance towards using television for language learning purposes, but the benefits should be clearly described beforehand. Learners ought to be told as well that perceiving improvement in their comprehension of the TV programmes is a very slow process, which will take time. Likewise, they need to be warned about the fact that some episodes will be easier or more difficult than others and that this should not discourage them.

Additionally, it is of vital importance that learners are at the appropriate level of proficiency to make sense of the input (Webb, 2015) and that the materials they are exposed to match their general proficiency level (i.e., such materials must be comprehensible). If the input is too challenging, they will not be able to understand it, and this will hinder vocabulary learning. One way to make input more accessible is the use of L1 or L2 subtitles. Bearing in mind that the present results were obtained with TV series with standard or bimodal subtitles, and, as subtitles tend to facilitate comprehension (Caimi, 2006), it is not delusive to think that learning gains would have been lower if the TV series had been watched without any textual support. In consequence, FL learners can be encouraged to use subtitles once they are proficient enough to process this textual support. Regarding the language of the subtitles, and based on the results of the thesis, it is advisable to use L1 subtitles with low-level learners, especially to foster the comprehension of the authentic input they will be exposed to. Although L2 subtitles were not used in Grade 6, participants' L1 reading speed was below the average speed with which subtitles are normally presented, which leads us to believe that using L2 subtitles at this level would certainly be counter-productive rather than facilitative. However, further research with young learners and different viewing options should be conducted. In contrast, in Grade 10 and university, it would be recommendable to use bimodal subtitling since the proficiency level is higher and L2 subtitles will grant more exposure to TL input, which might also foster the learning of word forms and may facilitate the linking of form and meaning. L1 subtitles could also be used at these levels if a very high degree of comprehension of FL television is the goal of the intervention.

Furthermore, regular use of FL television in the classroom context is recommended, as opposed to a single session. If a pedagogical programme needs to be designed from scratch, it is worth remembering that more benefits were reported towards the end of the academic year so an extended period over several years can be more appropriate. Hence, the emphasis should be on using television or videos on a regular basis, combined with other forms of instruction. However, it is very important to bear in mind that extensive viewing cannot replace formal vocabulary instruction. It can be regarded as a complement to enhance and reinforce language acquisition, but the benefits of curricular formal instruction cannot be neglected.

In sum, this thesis has shown the potential and the limitations of using FL television in the language classroom. It has revealed that additional exposure to TV series leads to more vocabulary learning than more traditional forms of instruction at certain proficiency levels, although focused learning is also necessary and exposure to TV series does not contribute to a steady increase of content comprehension after just several viewing sessions. If a similar intervention was to be implemented in other contexts, it is important to bear in mind that it ought to be accompanied by a set of focused tasks, that it should be sustained over time –and not limited to very few sessions–, that multimodal input should be presented with textual support whenever possible (especially in low and intermediate students), and that this should be in the learners' L1 in the case of low-level learners and, ideally, in the L2 with more advanced learners. Bearing all this in mind, it is hoped that the in-class learning experience will guide more out-of-class viewing (Webb, 2015), in which more vocabulary might be learned, and higher comprehension levels may be achieved.

9.3. Limitations of the study

It should be acknowledged that the present dissertation has some limitations, especially in relation to the study design. The fact that the study was longitudinal and set in a classroom context also caused several problems which do not tend to occur in other experimental settings. The selection of the input materials was also very challenging.

First of all, one of the major drawbacks is that there was not a comparison condition in which students watched the videos without subtitles. This group would have worked as an additional control condition and we would have been able to assess to what extent textual support was necessary for learning the TWs and for understanding the input participants were exposed to. Although the idea was considered, adding a third group was not possible in practical terms because there were not enough students available in the educational institutions where the study was conducted. Moreover, the main objective of this dissertation was to study how multimodal input (operationalized here as video viewing) affected (or not) language learning, and hence the central focus was not on the presence of subtitles, but on the use and effects of subtitled TV viewing in language classes at several proficiency levels. Subtitles were included to facilitate the understanding of the input by the EGs, which we assumed would have been problematic without them.

In addition, the inclusion of the vocabulary post-task may have affected the viewing experience as a whole: learners did not receive feedback on this task (it was not corrected in the classroom), so they could see it as an immediate post-test and made an extra effort to remember those words. They knew as well, after some sessions, that the TWs to which they were introduced with the pre-task would be asked again at the end of each viewing session. As a consequence, it is very likely that these words were the participants' main focus of attention, also when watching the TV episodes, and this could have prevented them from learning others. However, as all groups went through the same process (except for the EGs' extra exposure to the target vocabulary in the TV series), the effects of the processing of the TWs should logically have been the same for both experimental conditions.

Another problem we encountered in university contexts was that, as most courses at this proficiency level we had access to were not longer than one term, it was not possible to follow university participants throughout one academic year. Consequently, it would be very revealing to trace a university group's evolution throughout a longer period of time to see whether significant differences, which were not found in the present study, arose with more exposure to multimodal input (as it was the case in Grade 6 and Grade 10 towards the end of the intervention).

The fact that the study was longitudinal allowed us to trace participants' evolution and see the effects of sustained exposure to multimodal input, although two limitations should be pointed out. First, the test-training effect may have affected the scores on the post-tests, since learners were already familiar with the format and the instructions. Nevertheless, this usually happens with longitudinal studies assessing for language gains and using the same instrument as the pre- and post-test on several occasions. Second, as there were many different viewing and testing sessions throughout the academic year, some participants did not complete all the tests and tasks and the final sample of participants was reduced. This was thoroughly controlled for in the analysis, even if the final number of participants was sometimes limited, especially to conduct some of the statistical tests (e.g., the RM ANOVAs including the 22 or 24 comprehension tests); that is why alternative tests were performed each term. In addition, when participants were divided into the two conditions (experimental and control), the resulting number of students per group was sometimes small. Hence, a replication with a larger pool of participants would be needed to confirm or reject the present findings.

Even if classroom-based research is really needed in SLA (Lightbown, 2000), the fact that the study was set in a classroom context also entails some limitations. First, the viewing sessions were limited in number and they were all 60 minutes long. Ideally, longer TV series (e.g., lasting 40 minutes) would have been selected so as to expose learners to larger amounts of multimodal input, but they would not have fit in one hour, since other activities had to be done as well. Besides, it was impossible to complete the eight viewing sessions at T2 and T3 in Grade 10. Although efforts were made to make up for these sessions, it was not possible due to school holidays. Last, the same teacher was not responsible for the two experimental conditions in Grades 6 and 10. Even if all teachers involved in the intervention were trained and were given detailed instructions, only the first session was monitored. The researcher was not present in the rest (except in the administration of the proficiency, pre- and post-tests), so there may have been some minor differences in the way each teacher carried out the viewing sessions (for instance, by putting more or less emphasis on the vocabulary pre-task).

In connection with this, attendance was compulsory in Grades 6 and 10, so some students who might not be enthusiastic about the intervention had to attend the sessions anyway. This could have demotivated them and they could have consequently paid less attention to the activities. Similarly, students in the CGs may have felt a bit demotivated too because they knew that their peers were watching a TV series during class time and they wondered why they were not doing the same. As TV viewing is usually regarded as a fun experience, the scores of the CGs could have been affected by participants' low degree of motivation. Efforts were made to keep both conditions entertained and improve their motivation (e.g., by changing the TV series after several episodes if they asked for it).

Although much care was taken to choose the materials used in the intervention, working with authentic materials also poses some challenges and entails some limitations that need to be acknowledged. In this respect, in order to maintain learners entertained and willing to continue participating in the intervention, different TV series had to be used. It would have been desirable to have used the same series throughout the whole academic year, but we feared that students would get tired or would lose interest in watching the same TV series for nine months. As participants themselves verbalised their preference for a change of the series (this happened at the end of Term 2 in Grade 6 and at the end of Term 1 in Grade 10), we thought it would be better to meet their demands in order to foster their motivation and willingness to participate in the experiment. In such a long study, efforts were made so that students were directly involved and encouraged to learn during the viewing sessions. Despite not using the same series throughout the year, the two used at each of the levels have been shown to be comparable and of the same difficulty.

Regarding TWs' selection, they were chosen because they were thought to be unknown to learners and because they appear at least twice in the episode in which they were target. Nevertheless, although the TW sample was varied (e.g., different parts of speech, different frequencies, etc.), we were constrained by the vocabulary used in the episode and what learners could already know. That is one of the main problems of working with real input authentic materials which cannot be modified. We were very much restricted by the TV series themselves and, on very few occasions, we had to choose words that only appeared twice or that had some similarities with the Catalan / Spanish equivalent. However, this was balanced out by the fact that learners were tested on a total of 110-120 TWs and that these were of diverse nature: concrete, abstract, mostly non-cognates, frequent in the episodes, from a variety of grammatical categories, etc.

Finally, it is important to note that, when selecting participants from different proficiency levels, we inevitably have a confound between participants' biological age and their FL development. As already pointed out, participants in Grade 6 were younger and significantly less proficient than those in Grade 10. In turn, university learners were older and significantly more proficient than high-school students. It is then reasonable to think that the results of the thesis are not only influenced by participants' proficiency level, but also by their age: it was not possible in this study to select participants from different proficiency levels at the same age. Nevertheless, this situation mirrors what we actually find in the Catalan educational system (where participants are grouped in schools according to the year they were born, with older learners being more proficient than their younger peers) and it thus increases the ecological validity of the study (Spada, 2005).

9.4. Original contribution of the research

In spite of the weaknesses presented, the doctoral dissertation provides the field with original information on learning with multimodal materials and reflects on theoretical aspects related to this type of learning, which will be explained in this section. To start with, this dissertation puts different theories (DCT, CTML and CLT) to the test in a classroom context and it is one of the first that contributes to explain the role of their premises in FL learning. The thesis also assesses the implications of these theories in the long run, after sustained exposure to multimodal input and for different populations, revealing some findings that previous research did not identify (e.g., the DCT and CTML seem to be applicable to low-level EFL learners watching subtitled TV series only after considerable exposure to multimodal input).

Up to now, research has mainly focused on the usefulness of different subtitling modes or on the effectiveness of captions mostly in undergraduate students. Besides, the two studies which included a control condition with no exposure to video input (Peters & Webb, 2018; Rodgers, 2013) did not deliver any vocabulary pre-teaching but studied incidental learning. In this dissertation, the two experimental conditions were pre-taught the target vocabulary, but the EGs were additionally exposed to multimodal input: we did not compare test taking to incidental learning from video viewing, but we have been able to isolate the effect of additional video input exposure and analysed it thoroughly. Thus, this pedagogically-oriented study can be useful to evaluate extensive viewing as a complement to explicit instruction in classroom settings.

The present study also focuses on vocabulary retention, a factor that has not been much explored in relation to TV viewing. To the best of our knowledge, only very few studies on multimodal input included a delayed test in their design, but this was administered one or two weeks after the intervention. There are no instances of recent research including a delayed test some months after the experiment, to be able to study the long-term retention of the vocabulary learned through TV viewing. Besides, we used a more fine-grained measure of vocabulary retention (i.e., exclusively taking into account the knowledge actually learned during the intervention), which has also contributed to see that vocabulary long-term retention does not significantly vary when learned through curricular instruction or through additionally watching FL television. We think that studying long-term effects is necessary as an initial diagnosis, so as to see how we can best help our learners to make the most of the treatments.

In relation to the participants, and despite the problems outlined in the previous section, the study was carried out in intact language classes. In SLA, a field whose aim is to study and facilitate the acquisition of languages to populations of all kinds, it is of the utmost importance to conduct research on programmes and interventions that can actually be implemented in L1, L2 or FL classes, thus increasing ecological validity of the research (Spada, 2005).

In addition, we have also worked with primary school learners, with a beginner level of the TL, an under-researched population in SLA and in research on multimodal input (Montero Perez et al., 2013; Plonsky, 2015). Hence, the dissertation can be seen as a first attempt to know more about regular TV viewing and FL learning in children with low levels of English. The thesis is also the first to set up a pedagogical intervention using TV series and especially tailored materials at three different proficiency levels (primary school, high school and university), and the first as well to compare vocabulary learning and content comprehension across these levels.

In relation to the duration of the study, and as was acknowledged in the Literature Review, there was, and still is, a lack of longitudinal research on television viewing and language learning. The vast majority of research on multimodal input to date exposes learners to small amounts of such an input, and only Rodgers (2013), Frumuselu (2015) and Zarei (2009) included long-term exposure to video input. However, the three worked with adult university learners. This present dissertation is the first to expose younger and less proficient students to TV viewing for a whole academic year in real language classes. It also includes larger quantities of multimodal input (over eight hours in total and more than twenty episodes of television series) and it traces the evolution of participants' vocabulary learning and content comprehension throughout nine months. Thanks to this, we have been able to see the effects of extensive video viewing on the two areas under study in the longer term.

Last but not least, another contribution of this dissertation is the vocabulary test used, also taking into account aural word recognition. Most studies to date have used recognition tests, in which learners had to select the most appropriate form or meaning of the target vocabulary, but very few tested productive knowledge of the language. By using written form recall and meaning translation tests, the dissertation considers different aspects of vocabulary learning and takes into account the input students had been exposed to (oral and written).

9.5. Further research

There are many facets of language acquisition that this dissertation has not been able to analyse and that may have been enhanced by repeated video viewing. For example, this thesis has looked at vocabulary and content comprehension, but it would be very informative to see the effects of sustained exposure to TV viewing on syntax and grammar learning, pronunciation development or segmentation abilities.

Furthermore, no longitudinal studies have compared the effects of focused versus non-focused instruction (i.e., learning with or without any additional linguistic tasks) on vocabulary acquisition and retention through TV viewing. Hence, it would be very useful to see which gains could be achieved using different approaches and which is the best way to make the most out of the TV viewing experience (Pujadas & Muñoz, 2019). The same applies to the subtitling language. In this dissertation, we have used two types of subtitles (L1 in primary school and L2 in high school and university), following recommendations of previous research, but we have not compared the effectiveness of one versus the other by adding an extra group at each level. Hence, looking at the effects of the subtitling language in a study on TV viewing prolonged in time would be very revealing (Casulleras, forthcoming). The same holds true for other captioning techniques such as keyword, highlighted keyword or glossed captioning.

Another idea for future research has to do with the type of words students may learn from audiovisual materials. In connection with this, Peters and Webb (2018) found that words occurring more often are better learned and that "there is a facilitative effect for cognates" (p. 574), confirming what previous research has often suggested. However, it would be interesting to see whether this would be confirmed in a longitudinal study such as the one designed for this dissertation. Most importantly, it would be good to know whether some kind of words (e.g., nouns vs. verbs, concrete vs. abstract, etc.) can be better learned than others when exposed to multimodal input for a long period of time. In the same line, analysing whether imagery contributes to the learning of the target vocabulary (Rodgers, 2018) would also shed light on which videos could provide more chances for vocabulary learning.

Besides, with a higher number of participants in each of the groups, it would be interesting to conduct an intragroup variability analysis to see if more proficient students in the same class are also learning more vocabulary and achieving higher levels of content comprehension or, in contrast, if exposure to TV series better helps poor learners to develop their FL skills. Linking the results of this doctoral dissertation to individual differences such as language aptitude, WM or inhibitory control would also help to understand how we learn from TV viewing (see Suárez & Gesa, 2019) and would contribute to the adaptation of the intervention for different populations.

Finally, other ideas for further research include: (1) extending the intervention at the university level, this time during one academic year, to be able to trace participants' evolution and compare their lexical gains with other grades; (2) replicating the study with a larger pool of participants; (3) conducting a similar study with advanced learners (e.g., English majors) exposed to video viewing without textual support, to see the effects of captioning on the learning of vocabulary and on the evolution of content comprehension; or (4) studying low-level learners with different types of real TV series or cartoons.

9.6. Concluding remarks

The main objective of this PhD thesis has been exploring language learning when viewing L1 or L2 subtitled TV series. It adds some results on the scarce longitudinal research on the topic at different proficiency levels, and its findings provide insight into: (1) vocabulary learning and long-term retention from viewing subtitled FL television along with, and as compared to, formal curricular instruction, and (2) content comprehension of FL subtitled TV series after regular exposure to different episodes. We thus hope that this dissertation will promote further research on multimodal input and language learning in different settings and with different populations. Finally, we do fervently hope that the in-class learning from video viewing that participants in this dissertation experienced will encourage them to continue watching FL television outside the classroom, which will in time enhance FL learning (Webb, 2015).

REFERENCES

Alexiou, T. (2015). Vocabulary uptake from *Peppa Pig*: A case study of preschool EFL learners in Greece. In C. Gitsaki, & T. Alexiou (Eds.), *Current issues in second / foreign language teaching and teacher development: Research and practice* (pp. 285-301). Newcastle upon Tyne: Cambridge Scholars Publishing.

Alipour, M., Gorjian, B., & Kouravand, L. G. (2012). The effects of pedagogical and authentic films on EFL learners' vocabulary learning: The role of subtitles. *Advances in Asian Social Sciences (AASS), 3*(4), 734-738.

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

Alloway, T. P., Williams, S., Jones, B., & Cochrane, F. (2014). Exploring the impact of television watching on vocabulary skills in toddlers. *Early Childhood Education Journal, 42*(5), 343-349. doi: 10.1007/s10643-013-0618-1

Almeida, P. A., & Costa, P. D. (2014). Foreign language acquisition: the role of subtitling. *Procedia - Social and Behavioral Sciences, 14*, 1234-1238. doi: 10.1016/j.sbspro.2014.05.212

Anderson, N. J. (2002). *The role of metacognition in second language teaching and learning. ERIC Digest*. Retrieved from ERIC database (EDO-FL-01-10): https://www.govinfo.gov/content/pkg/ERIC-ED463659/pdf/ERIC-ED463659.pdf

Atkins, P. W. B., & Baddeley, A. (1998). Working memory and distributed vocabulary learning. *Applied Psycholinguistics, 19*(4), 537-552. doi: 10.1017/S0142716400010353

Aurstad, L. M. G. (2013). *The role of subtitles in second language acquisition. An experimental study in the context of the Norwegian school system* (Master's thesis, Norwegian University of Science and Technology, Trondheim, Norway). Retrieved from: https://brage.bibsys.no/xmlui/handle/11250/243562

Austin, A., Barnard, J., & Hutcheon, N. (2016). *Media consumption forecasts 2016*. Retrieved from Zenith Optimedia: https://www.zenithmedia.com/

Baddeley, A. (1986). *Working memory*. Oxford: Oxford University Press.

Baddeley, A. (1992). Working memory. *Science, 255*(5044), 556-559. doi: 10.1126/science.1736359

Baddeley, A. (2000). The episodic buffer: a new component of working memory? *Trends in Cognitive Sciences, 4*(11), 417-423. doi: 10.1016/S1364-6613(00)01538-2

Bae, H. S. (2015). Metalinguistic awareness in vocabulary acquisition: Evidence from grade six Korean EFL learners. *Primary English Education, 21*(3), 5-22.

Baltova, I. (1999). *The effect of subtitled and staged video input on the learning and retention of content and vocabulary in a second language* (Doctoral dissertation, University of Toronto, Toronto, Canada). Retrieved from: https://tspace.library.utoronto.ca/

Barclay, S. (2018, March). *Understanding the role of systematic decay as part of vocabulary acquisition*. Paper presented at the 2018 American Association for Applied Linguistics (AAAL) Conference, Chicago, United States.

Barlovento Comunicación. (2018). *Análisis televisivo 2017*. Retrieved from Barlovento Comunicación: http://www.barloventocomunicacion.es/wp-content/uploads/2018/01/analisis-televisivo-2017-Barlovento.pdf

Becker, S. R., & Sturm, J. L. (2017). Effects of audiovisual media on L2 listening comprehension: A preliminary study in French. *CALICO Journal, 34*(2), 147-177. doi: 10.1558/cj.26754

Bensoussan, M., & Laufer, B. (1984). Lexical guessing in context in EFL reading comprehension. *Journal of Research in Reading, 7*(1), 15-31. doi: 10.1111/j.1467-9817.1984.tb00252.x

Bird, S. A., & Williams, J. N. (2002). The effect of bimodal input on implicit and explicit memory: An investigation into the benefits of within-language subtitling. *Applied Psycholinguistics, 23*(4), 509-533. doi: 10.1017.S0142716402004022

Bisson, M.-J., van Heuven, W. J. B., Conklin, K., & Tunney, R. J. (2014a). Processing of native and foreign language subtitles in films: An eye tracking study. *Applied Psycholinguistics, 35*(2), 399-418. doi: 10.1017/S0142716412000434

Bisson, M.-J., van Heuven, W. J. B., Conklin, K., & Tunney, R. J. (2014b). The role of repeated exposure to multimodal input in incidental acquisition of a foreign language vocabulary. *Language Learning, 64*(4), 855-877. doi: 10.1111/lang.12085

Borrás, I., & Lafayette, R. C. (1994). Effects of multimedia courseware subtitling on the speaking performance of college students of French. *The Modern Language Journal, 78*(1), 61-75. doi: 10.2307/329253

Bravo, M. C. C. (2008). *Putting the reader in the picture: Screen translation and foreign-language learning*. (Doctoral dissertation, Universitat Rovira i Virgili, Tarragona, Spain). Retrieved from: http://www.sub2learn.ie/downloads/tesicondhino.pdf

Brett, P. (1997). A comparative study on the effects of the use of multimedia on listening comprehension. *System, 25*(1), 39-53. doi: 10.1016/S0346-251X(96)00059-0

Brown, R., Waring, R., & Donkaewbua, S. (2008). Incidental vocabulary acquisition from reading, reading-while-listening, and listening to stories. *Reading in a Foreign Language, 20*(2), 136-163. Retrieved from: http://nflrc.hawaii.edu/rfl/October2008/brown/brown.pdf

Bruton, A., García López, M., & Esquiliche Mesa, R. (2011). Incidental L2 vocabulary learning: An impracticable term? *TESOL Quarterly, 45*(4), 759-768. doi: 10.5054/tq.2011.268061

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods, 41*(9), 977-990. doi: 10.3758/BRM.41.4.977

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods, 46*(3), 904-911. doi: 10.3758/s13428-013-0403-5.

Buck, G. (2001). *Assessing listening*. Cambridge: Cambridge University Press.

Bygate, M., & Samuda, V. (2005). Integrative planning through the use of task repetition. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 37-74). Amsterdam: John Benjamins.

Caimi, A. (2006). Audiovisual translation and language learning: The promotion of intralingual subtitles. *The Journal of Specialised Translation, 6*, 85-98. Retrieved from: https://www.jostrans.org/issue06/art_caimi.php

Cain, K., & Oakhill, J. (2006). Assessment matters: Issues in the measurement of reading comprehension. *British Journal of Educational Psychology, 76*(4), 697-708. doi: 10.1348/000709905X69807

Casulleras, M. (forthcoming). *The use of subtitled TV series in the EFL primary classroom* (Doctoral dissertation, Universitat de Barcelona, Barcelona, Spain).

Chafe, W. (1994). *Discourse, consciousness and time*. Chicago: University of Chicago Press.

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction, 8*(4), 293-332. doi: 10.1207/s1532690xci0804_2

Chang, A. C.-S. (2009). Gains to L2 listeners from reading while listening vs. listening only in comprehending short stories. *System, 37*, 652-663. doi: 10.1016/j.system.2009.09.009

Chang, A. C.-S. (2012). Gains to L2 learners from extensive listening: Listening development, vocabulary acquisition and perceptions of the intervention. *Hong Kong Journal of Applied Linguistics, 14*(1), 25-47.

Charles, T., & Trenkic, D. (2015). Speech segmentation in a SL: The role of bi-modal input. In Y. Gambier, A. Caimi, & C. Mariotti (Eds.), *Subtitles and language learning. Principles, strategies and practical experiences* (pp. 173-197). Bern: Peter Lang.

Chun, D. M., & Plass, J. L. (1996). Effects of multimedia annotations on vocabulary acquisition. *The Modern Language Journal*, *80*(2), 183-198. doi: 10.1111/j.1540-4781.1996.tb01159.x

Clark, J. M., & Paivio, A. (1991). Dual Coding Theory and education. *Educational Psychology Review, 3*(3), 149-210. doi: 10.1007/BF01320076

Cobb, T. (2007). Computing the vocabulary demands of L2 reading. *Language, Learning & Technology, 11*(3), 38.63. doi: 10125/44117

Cobb, T. (ongoing). VocabProfile (version 2) [Computer software]. Retrieved from: https://www.lextutor.ca/vp/comp/

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (second edition). Mahwah: Lawrence Erlbaum Associates.

Cokely, M., & Muñoz, C. (2019). Captioned video, vocabulary and visual prompts: An exploratory study. In C. Herrero, & I. Vanderschelden (Eds.), *Using film and media in the language classroom: Reflections on research-led teaching* (pp. 61-75). Bristol: Multilingual Matters.

Dafei, D. (2007). An exploration of the relationship between learner autonomy and English proficiency. *Asian EFL Journal*, *24*, 1-23. Retrieved from: http://asian-efl-journal.com/pta_Nov_07_dd.pdf

Dahlen, K., & Caldwell-Harris, C. (2013). Rehearsal and aptitude in foreign vocabulary learning. *The Modern Language Journal, 97*(4), 902-916. doi: 10.1111/j.1540-4781.2013.12045.x

Danan, M. (1992). Reversed subtitling and Dual Coding Theory: New directions for foreign language instruction. *Language Learning, 42*(4), 497-527. doi: 10.1111/j.1467-1770.1992.tb01042.x

Danan, M. (2004). Captioning and subtitling: Undervalued language learning strategies. *Meta: journal des traducteurs / Meta: translators' journal, 49*(1), 67-77. doi: 10.7202/009021ar

Danan, M. (2015). Subtitling as a language learning tool: Past findings, current applications, and future paths. In Y. Gambier, A. Caimi, & C. Mariotti (Eds.), *Subtitles and language learning. Principles, strategies and practical experiences* (pp. 41-61). Bern: Peter Lang.

Darmi, R. (2012). Designing tasks for the second language class: Is there a role for contextual support and cognitive challenge in second language learning? *Frontiers of Language and Teaching, 3*, 139-146.

Date (n.d.). In *Oxford English Dictionary*. Retrieved from: https://www.lexico.com/en/definition/date

David, L., Shapiro, G., West, H., Scheinman, A., Seinfeld, J., Berg, A., & Schaffer, J. (Producers). (1989). *Seinfeld* [Television series]. New York: National Broadcasting Company.

Davies, M. (2009). The 385+ million word *Corpus of Contemporary American English* (1990-2008+). Design, architecture, and linguistic insights. *International Journal of Corpus Linguistics, 14*(2), 159-190. doi: 10.1075/ijcl.14.2.02dav

Day, R. R., & Bamford, J. (1998). *Extensive reading in the second language classroom*. Cambridge: Cambridge University Press.

Day, R. R., & Park, J.-S. (2005). Developing reading comprehension questions. *Reading in a Foreign Language, 17*(1), 60-73. Retrieved from: http://nflrc.hawaii.edu/rfl/April2005/day/day.pdf

del Rincón, D., Arnal, J., Latorre, A., & Sans, A. (1995). *Técnicas de investigación en ciencias sociales*. Madrid: Dykinson.

Díaz Cintas, J. (2010). Subtitling. In Y. Gambier, & L. van Doorslaer (Eds.), *Handbook of translation studies. Volume 1* (pp. 344-349). Amsterdam: John Benjamins.

Dörnyei, Z. (2005). *The psychology of the language learner: Individual differences in second language acquisition*. Mahwah: Lawrence Erlbaum Associates.

Dunn, L. M., & Dunn, L. M. (1981). *The Peabody Picture Vocabulary Test* (second edition). Bloomington: NCS Pearson, Inc.

Dunn, L. M., & Dunn, D. M. (2007). *The Peabody Picture Vocabulary Test* (fourth edition). Bloomington: NCS Pearson, Inc.

d'Ydewalle, G., & De Bruycker, W. (2007). Eye movements of children and adults while reading television subtitles. *European Psychologist, 12*(3), 196-205. doi: 10.1027/1016-9040.12.3.196

d'Ydewalle, G., & Pavakanun, U. (1995). Acquisition of a second / foreign language by viewing a television program. In P. Winterhoff-Spurk (Ed.), *Psychology of media in Europe: The state of the art - perspectives for the future* (pp. 51-64). Opladen: Westdeutscher Verlag.

d'Ydewalle, G., & Pavakanun, U. (1997). Could enjoying a movie lead to language acquisition? In P. Winterhoff-Spurk, & T. H. A. van der Voort (Eds.), *New horizons in media psychology: Research cooperation and projects in Europe* (pp. 145-155). Wiesbaden: Springer Fachmedien.

d'Ydewalle, G., Praet, C., Verfaillie, K., & Van Rensbergen, J. (1991). Watching subtitled television: Automatic reading behavior. *Communication Research, 18*(5), 650-666. doi: 10.1177/009365091018005005

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. doi: 10.1023/A:1023202130625

Elley, W. B. (1989). Vocabulary acquisition from listening to stories. *Reading Research Quarterly, 24*(2), 147-189. doi: 10.2307/747863

Ellis, N. Personal interview. 21 October 2016.

Ellis, R. (1999). Factors in the incidental acquisition of second language vocabulary from oral input. In R. Ellis (Ed.), *Learning a second language through interaction* (pp. 35-61). Amsterdam: John Benjamins.

Ellis, R., & Barkhuizen, G. (2005). *Analysing learner language*. Oxford: Oxford University Press.

Etemadi, A. (2012). Effects of bimodal subtitling of English movies on content comprehension and vocabulary recognition. *International Journal of English Linguistics, 2*(1), 239-248. doi:10.5539/ijel.v2n1p239

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). Retrieved from: https://publications.europa.eu/en/publication-detail/-/publication/e4d5cbf4-a839-4a8a-81d0-7b19a22cc5ce

European Commission. (2017). *Media use in the European Union* (Standard Eurobarometer 88). Retrieved from: http://ec.europa.eu/commfrontoffice/publicopinion/index.cfm/Survey/getSurveyDetail/instruments/STANDARD/surveyKy/2143

Field, A. (2009). *Discovering statistics using SPSS* (third edition). London: SAGE Publications.

Flavell, J. H. (1979). Metacognition and cognitive monitoring. A new area of cognitive-development inquiry. *American Psychologist, 34*(10), 906-911. doi: 10.1037/0003-066X.34.10.906

Frumuselu, A. C. (2015). *Subtitled television series inside the EFL classroom: Long-term effects upon colloquial language learning and oral production*. (Doctoral dissertation, Universitat Rovira i Virgili, Tarragona, Spain). Retrieved from: https://www.tesisenred.net/handle/10803/378642

Frumuselu, A. C., De Maeyer, S., Donche, V., & Gutiérrez Colon Plana, M. 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. doi: 10.1016/j.linged.2015.10.001

Fundación Telefónica (2016). *La sociedad de la información en España 2016* (Report No. 36). Barcelona: Ariel.

Galí, A. (1984). *La mesura objectiva del treball escolar* (second edition). Vic: Eumo Editorial / Diputació de Barcelona.

Galimberti, V. (2016). *Watching TV series with L1 or L2 subtitles: A study on comprehension and vocabulary acquisition with Italian beginner learners of English.* (Master's thesis, Universitat de Barcelona, Barcelona, Spain).

Gambier, Y. (2015). Subtitles and Language Learning (SLL): Theoretical background. In Y. Gambier, A. Caimi, & C. Mariotti (Eds.), *Subtitles and language learning. Principles, strategies and practical experiences* (pp. 63-82). Bern: Peter Lang.

García, B. (2017). Bilingual subtitles for second-language acquisition and application to engineering education as learning pills. *Computer applications in engineering education, 25*(3), 468-479. doi: 10.1002/cae.21814

Garnier, M. (2014). Intentional vocabulary learning from watching DVDs with subtitles: A case study of an 'average' learner of French. *International Journal of Research Studies in Language Learning, 3*(1), 21-32. doi: 10.5861/ijrsll.2013.521

Garza, T. J. (1991). Evaluating the use of captioned video materials in advanced foreign language learning. *Foreign Language Annals, 24*(3), 239-258. doi: 10.1111/j.1944-9720.1991.tb00469.x

Gass, S. (1999). Discussion: Incidental vocabulary learning. *Studies in Second Language Acquisition, 21*(2), 319-333. Retrieved from: https://www.cambridge.org/core/journals/studies-in-second-language-acquisition

Gilmore, A. (2007). Authentic materials and authenticity in foreign language learning. *Language Teaching, 40*(2), 97-118. doi:10.1017/S0261444807004144

Granena, G., Muñoz, C., & Tragant, E. (2015). L1 reading factors in extensive L2 reading-while-listening instruction. *System, 55*, 86-99. doi: 10.1016/j.system.2015.09.005

Greenwald, T. J., Murrieta, P., Cheung, V., & Montanio, B. (Producers). (2007). *Wizards of Waverly Place* [Television series]. Bunbark: Disney Channel.

Gu, P. Y. (2003). Vocabulary learning in a second language: Person, task, context and strategies. *TESL-EJ, 7*(2), 1-25. Retrieved from: http://tesl-ej.org/ej26/a4.html

Gu, P. Y. (2013). Second language vocabulary. In J. Hattie, & E. M. Anderson (Eds.), *International guide to student achievement* (pp. 307-309). New York: Routledge.

Guillory, H. G. (1998). The effects of keyword captions to authentic French video on learner comprehension. *CALICO Journal, 15*(1-3), 89-108. doi: 10.1558/cj.v15i1-3.89-108

Haahr, M., & Haahr, S. (1998). RANDOM.ORG - True Random Number Service. Retrieved from https://www.random.org/.

Hayati, A., & Mohmedi, F. (2011). The effect of films with and without subtitles on listening comprehension of EFL learners. *British Journal of Educational Technology, 42*(1), 181-192. doi: 10.1111/j.1467-8535.2009.01004.x

Herron, C. (1994). An investigation of the effectiveness of using an advance organizer to introduce video in the foreign language classroom. *The Modern Language Journal, 78*(2), 190-198. doi: 10.2307/329009

Holec, H. (1981). *Autonomy and foreign language learning*. Oxford: Pergamon Press.

Holobow, N. E., Lambert, W. E., & Sayegh, L. (1984). Pairing script and dialogue: Combinations that show promise for second or foreign language learning. *Language Learning, 34*(4), 59-76. doi: 10.1111/j.1467-1770.1984.tb00352.x

Hopkins, C. D., & Antes, R. L. (1978). *Classroom measurement and evaluation*. Itasca: Peacock Publishers.

Horst, M., Cobb, T., & Meara, P. M. (1998). Beyond a clockwork orange: Acquiring second language vocabulary through reading. *Reading in a Foreign Language, 11*(2), 207-223. Retrieved from: http://nflrc.hawaii.edu/rfl/PastIssues/rfl112horst.pdf

Hoven, D. (1999). A model for listening and viewing comprehension in multimedia environments. *Language, Learning & Technology, 3*(1), 73-90. doi: 10125/25058

Hsu, W. (2013). The effects of audiovisual support on EFL learners' productive vocabulary. *ReCALL, 26*(1), 62-79. doi: 10.1017/S0958344013000220

Hu, M., & Nation, I. S. P. (2000). Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language, 13*(1), 403-430. Retrieved from: http://nflrc.hawaii.edu/rfl/PastIssues/rfl131hsuehchao.pdf

Huang, H.-C., & Eskey, D. E. (1999). The effects of closed-captioned television on the listening comprehension of intermediate English as a second language (ESL) students. *Journal of Educational Technology Systems, 28*(1), 75-96. doi: 10.2190/RG06-LYWB-216Y-R27G

Huckin, T., & Coady, J. (1999). Incidental vocabulary acquisition in a second language. A review. *Studies in Second Language Acquisition, 21*(2), 181-193. Retrieved from: https://www.cambridge.org/core/journals/studies-in-second-language-acquisition

Hulstijn, J. H. (2001). Intentional and incidental second language vocabulary learning: a reappraisal of elaboration, rehearsal and automaticity. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 258-286). Cambridge: Cambridge University Press.

Hulstijn, J. H. (2003). Incidental and intentional learning. In C. J. Doughty, & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 349–381). Oxford: Blackwell.

Hulstijn, J. H. (2013). Incidental learning in second language acquisition. In C. A. Chapelle (Ed.), *The encyclopedia of Applied Linguistics* (pp. 2,632-2,640). Chichester: Wiley-Blackwell.

Jacobs, M., Kendall, D., Young, B., Busgang, H., Blutman, M., & Tischler, B. (Producers). (1993). *Boy Meets World* [Television series]. New York: American Broadcasting Company.

Jamet, E., & Bohec, O. (2007). The effect of redundant text in multimedia instruction. *Contemporary Educational Psychology, 32*(4), 588–598. doi: 10.1016/j.cedpsych.2006.07.001

Jones, L. C., & Plass, J. L. (2002). Supporting listening comprehension and vocabulary acquisition in French with multimedia annotations. *The Modern Language Journal, 86*(4), 546-561. doi: 10.1111/1540-4781.00160

Kallis, D., Dreayer, I., Eells, P., & Geoghan, J. (Producers). (2005). *The Suite Life of Zack and Cody* [Television series]. Burbank: Disney Channel.

Kalyuga, S. (2005). Prior knowledge principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (first edition) (pp. 325-338). Cambridge: Cambridge University Press.

Kalyuga, S. (2011). Cognitive load theory: Implications for affective computing. In P. M. McCarthy, & R. C. Murray (Eds.), *Proceedings of the 24th International Florida Artificial Intelligence Research Society Conference* (pp. 105-110). Menlo Park: Association for the Advancement of Artificial Intelligence (AAAI).

Kalyuga, S., Chandler, P., Tuovinen, J., & Sweller, J. (2001). When problem solving is superior to studying worked examples. *Journal of Educational Psychology, 93*(3), 579–588. doi: 10.1037/0022-0663.93.3.579

Karakas, A., & Saricoban, A. (2012). The impact of watching subtitled animated cartoons on incidental vocabulary learning of ELT students. *Teaching English with Technology, 12*(4), 3-15. Retrieved from: http://cejsh.icm.edu.pl/cejsh/element/bwmeta1.element.desklight-b682ecc0-01d2-4e09-a861-c94d69984460

Kim, Y., Crossley, S., Jung, Y., Kyle, K., & Kang, S. (2018). The effects of task repetition and task complexity on L2 lexicon use. In M. Bygate (Ed.), *Learning language through task repetition* (pp. 75-96). Amsterdam: John Benjamins.

Koblin, J. (2016, June 30). How much do we love TV? Let us count the ways. *The New York Times*. Retrieved from: https://www.nytimes.com

Koolstra, C. M., & Beentjes, J. W. J. (1999). Children's vocabulary acquisition in a foreign language through watching subtitled television programs at home. *Educational technology research and development, 47*(1), 51-60. doi: 10.1007/BF02299476

Koolstra, C. M., Peeters, A. L., & Spinhof, H. (2002). The pros and cons of dubbing and subtitling. *European Journal of Communication, 17*(3), 325-354. doi: 10.1177/0267323102017003694

Kovacs, G., & Miller, R. C. (2014). Smart subtitles for vocabulary learning. In *Proceedings of the 2014 SIGCHI Conference on Human Factors in Computing Systems* (pp. 853-862). New York: Association for Computing Machinery.

Krashen, S. D. (1982). *Principles and practice in second language acquisition*. Oxford: Pergamon.

Krashen, S. D. (1985). *The Input Hypothesis: Issues and implications*. London and New York: Longman.

Kruger, J.-L., Hefer-Jordaan, E., & Matthew, G. (2013). Measuring the impact of subtitles on cognitive load: Eye tracking and dynamic audiovisual texts. In *Proceedings of the 2013 Conference on Eye Tracking South Africa* (pp. 62-66). New York: Association for Computing Machinery.

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

Kvitnes, I. E. N. (2013). *Subtitles in the second language classroom. An experimental study with Norwegian learners of English* (Master's thesis, Norwegian University of Science and Technology, Trondheim, Norway). Retrieved from: https://brage.bibsys.no/xmlui/handle/11250/243554

Landerl, K., & Wimmer, H. (2008). Development of word reading fluency and spelling in a consistent orthography: An 8-year follow-up. *Journal of Educational Psychology, 100*(1), 150-161. doi: 10.1037/0022-0663.100.1.150

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

Laufer, B. (1998). The development of passive and active vocabulary in a second language: Same or different? *Applied Linguistics, 19*(2), 255-271. doi: 10.1093/applin/19.2.255

Laufer, B. (2006). Comparing focus on form and focus on formS in second-language vocabulary learning. *The Canadian Modern Language Review / La Revue canadienne des langues vivantes, 63*(1), 149-166. doi: 10.1353/cml.2006.0047

Laufer, B., & Ravenhorst-Kalovski, G. C. (2010). Lexical threshold revisited: Lexical text coverage, learners' vocabulary size and reading comprehension. *Reading in a Foreign Language, 22*(1), 15-30. Retrieved from: https://files.eric.ed.gov/fulltext/EJ887873.pdf

Lavaur, J.-M., & Bairstow, D. (2011). Languages on the screen: Is film comprehension related to the viewers' fluency level and to the language of the subtitles? *International Journal of Psychology, 46*(6), 455-462. doi: 10.1080/00207594.2011.565343

Lee, M., Roskos, B., & Ewoldsen, D. R. (2013). The impact of subtitles on comprehension of narrative film. *Media Psychology, 16*(4), 412-440. doi: 10.1080/15213269.2013.826119

Lertola, J. (2012). The effect of the subtitling task on vocabulary learning. In A. Pym, & D. Orrego-Carmona (Eds.), *Translation Research Projects 4* (pp. 61-70). Tarragona: Intercultural Studies Group.

Li, S. (2010). The effectiveness of corrective feedback in SLA: A meta-analysis. *Language Learning, 60*(2), 309-365. doi: 10.1111/j.1467-9922.2010.00561.x

Lightbown, P. M. (2000). Anniversary article. Classroom SLA research and second language teaching. *Applied Linguistics, 21*(4), 431-462. doi: 10.1093/applin/21.4.431

Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In W. C. Ritchie, & T. K. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 413-468). San Diego: Academic Press.

Long, M. H. (1998). Focus on form in task-based language teaching. *University of Hawai'i Working Papers in ESL, 16*(1), 35-49. Retrieved from: https://core.ac.uk/download/pdf/77238837.pdf

Malone, J. (2018). Incidental vocabulary learning in SLA: Effects of frequency, aural enhancement, and working memory. *Studies in Second Language Acquisition, 40*(3), 651-675. doi: 10.1017/S0272263117000341

Maneshi, N. (2017). *Incidental vocabulary learning through listening to songs* (Master's thesis, University of Western Ontario, London, Canada). Retrieved from: https://ir.lib.uwo.ca/cgi/viewcontent.cgi?referer=https://www.google.com/&httpsredir=1&article=6701&context=etd

Mariotti, C. (2015). 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. Principles, strategies and practical experiences* (pp. 83-103). Bern: Peter Lang.

Markham, P. L. (1999). Captioned videotapes and second-language listening word recognition. *Foreign Language Annals, 32*(3), 321-328. doi: 10.1111/j.1944-9720.1999.tb01344.x

Markham, P. L., Peter, L. A., & McCarthy, T. J. (2001). The effects of native language vs. target language captions on foreign language students' DVD video comprehension. *Foreign Language Annals, 34*(5), 439-445. doi: 10.1111/j.1944-9720.2001.tb02083.x

Mayer, R. E. (2002). Cognitive theory and the design of multimedia instruction: An example of the two-way street between cognition and instruction. *New Directions for Teaching and Learning, 89*, 55-71. doi: 10.1002/tl.47

Mayer, R. E. (2009). *Multimedia learning* (second edition). Cambridge: Cambridge University Press.

Mayer, R. E. (2014). Cognitive Theory of Multimedia Learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (second edition) (pp. 43-71). Cambridge: Cambridge University Press.

Mayer, R. E., & Fiorella, L. (2014). Principles for reducing extraneous processing in multimedia learning: Coherence, signaling, redundancy, spatial contiguity, and temporal contiguity techniques. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (second edition) (pp. 279-315). Cambridge: Cambridge University Press.

Mayer, R. E., Heiser, H., & Lonn, S. (2001). Cognitive constraints on multimedia learning: When presenting more material results in less understanding. *Journal of Educational Psychology, 93*(1), 187-198. doi: 10 1037i/0022-0663 93.1 187

Mayer, R. E., Lee, H., & Peebles, A. (2014). Multimedia learning in a second language: A cognitive load perspective. *Applied Cognitive Psychology, 28*(5), 653-660. doi: 10.1002/acp.3050

Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist, 38*(1), 43–52. doi: 10.1207/S15326985EP3801_6

McNamara, T. (2000). *Language Testing.* Oxford: Oxford University Press.

Meara, P. M. (1994). LLEX: Lingua vocabulary tests (version 1.4) [Computer software]. Swansea: Centre for Applied Language Studies, University of Wales.

Meara, P. M. (2005). X_Lex: the Swansea vocabulary levels test (version 2.05) [Computer software]. Swansea: Centre for Applied Language Studies, University of Wales.

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

Meara, P. M., & Miralpeix, I. (2006). Y_Lex: the Swansea advanced vocabulary levels test (version 2.05). [Computer software]. Swansea: Centre for Applied Language Studies, University of Wales.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*(2), 81-97. doi: 10.1037/h0043158

Miller, G. A., & Chapman, J. P. (2001). Misunderstanding analysis of covariance. *Journal of Abnormal Psychology, 110*(1), 40-48. doi: 10.1037//0021-843X.110.1.40

Milton, J. (2008). Vocabulary uptake from informal learning tasks. *Language Learning Journal, 36*(2), 227-237. doi: 10.1080/09571730802390742

Miralpeix, I. (2012, March). *X_Lex and Y_Lex: A validation study*. Paper presented at the 22nd Lexical Studies Conference, Newtown, United Kingdom.

Miralpeix, I., & Muñoz, C. (2018). Receptive vocabulary size and its relationship to EFL language skills. *IRAL - International Review of Applied Linguistics in Language Teaching, 56*(1), 1-24. doi: 10.1515/iral-2017-0016

Mitterer, H., & McQueen, J. M. (2009). Foreign subtitles help but native-language subtitles harm foreign speech perception. *PLoS ONE, 4*(11), e7785. doi: 10.1371/journal.pone.0007785

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. doi: 10125/44357

Montero Perez, M., Peters, E., & Desmet, P. (2013). Is less more? Effectiveness and perceived usefulness of keyword and full captioned video for L2 listening comprehension. *ReCALL, 26*(1), 21-43. doi: 10.1017/S0958344013000256

Montero Perez, M., Peters, E., & Desmet, P. (2015). Enhancing vocabulary learning through captioned video: An eye-tracking study. *The Modern Language Journal, 99*(2), 308-328. doi: 10.1111/modl.12215

Montero Perez, M., Peters, E., & Desmet, P. (2018). Vocabulary learning through viewing video: the effect of two enhancement techniques. *Computer Assisted Language Learning, 31*(1-2), 1-26. doi: 10.1080/09588221.2017.1375960

Montero Perez, M., Van Den Noortgate, W., & Desmet, P. (2013). Captioned video for L2 listening and vocabulary learning: A meta-analysis. *System, 41*(3), 720-739. doi: 10.1016/j.system.2013.07.013

Moreno, R., & Mayer, R. E. (2002). Verbal redundancy in multimedia learning: When reading helps listening. *Journal of Educational Psychology, 94*(1), 156-163. doi: 10.1037//0022-0663.94.1.156

Morton, J. (1969). Interaction of information in word recognition. *Psychological Review, 76*(2), 165-178. doi: 10.1037/h0027366

Mousavi, S. Y., Low, R., & Sweller, J. (1995). Reducing cognitive load by mixing auditory and visual presentation modes. *Journal of Educational Psychology, 87*(2), 319–334. doi: 10.1037/0022-0663.87.2.319

Muñoz, C. (Ed.) (2006). *Age and the rate of foreign language learning*. Clevedon: Multilingual Matters.

Muñoz, C. (2008). Symmetries and asymmetries of age effects in naturalistic and instructed L2 learning. *Applied Linguistics, 29*(4), 578-596. doi: 10.1093/applin/amm056

Nagira, Y. (2011). Vocabulary learning through captions. *International Proceedings of Economics Development and Research, 26*, 95-99. Retrieved from: http://www.ipedr.com/vol26/20-ICLLL%202011-L00049.pdf

Nagy, W. E., Herman, P. A., & Anderson, R. C. (1985). Learning words from context. *Reading Research Quarterly, 20*(2), 233-253. doi: 10.2307/747758

Nakata, T. (2014). Effects of feedback timing on second language vocabulary learning: Does delaying feedback increase learning? *Language Teaching Research, 19*(4), 416-434. doi: 10.1177/1362168814541721

Nation, I. S. P. (1983). Testing and teaching vocabulary. *Guidelines, 5*(1), 12-25. Retrieved from: https://www.victoria.ac.nz/lals/about/staff/publications/paul-nation/1983-Testing-and-teaching.pdf

Nation, I. S. P. (1990). *Teaching and learning vocabulary.* Boston: Heinle & Heinle Publishers.

Nation, I. S. P. (2001). *Learning vocabulary in another language* (first edition). Cambridge: Cambridge University Press.

Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *The Canadian Modern Language Review / La Revue canadienne des langues vivantes, 63*(1), 59-82. doi: 10.3138/cmlr.63.1.59

Nation, I. S. P. (2007). The four strands. *Innovation in Language Learning and Teaching, 1*(1), 2-13. doi: 10.2167/illt039.0

Nation, I. S. P. (2012, September 17). The BNC/COCA word family lists. [PDF]. Retrieved from: https://www.victoria.ac.nz/lals/about/staff/publications/paul-nation/Information-on-the-BNC_COCA-word-family-lists.pdf

Nation, I. S. P. (2013). *Learning vocabulary in another language* (second edition). Cambridge: Cambridge University Press.

Nation, I. S. P., & Heatley, A. (2002). RANGE: A program for the analysis of vocabulary in texts [Computer software]. Retrieved from: https://www.victoria.ac.nz/lals/about/staff/paul-nation#vocab-programs

Nation, I. S. P., & Ming-Tzu, K. W. (1999). Graded readers and vocabulary. *Reading in a Foreign Language, 12*(2), 355-380. Retrieved from: http://nflrc.hawaii.edu/rfl/PastIssues/rfl122nation.pdf

Nation, I. S. P., & Webb, S. (2011). *Researching and analyzing vocabulary*. Boston: Heinle, Cengage Learning.

Neuman, S. B., & Koskinen, P. (1992). Captioned television as comprehensible input: Effects of incidental word learning from context for language minority students. *Reading Research Quarterly, 27*(1), 94-106. doi: 10.2307/747835

New, B., Brysbaert, M., Veronis, J., & Pallier, C. (2007). The use of film subtitles to estimate word frequencies. *Applied Psycholinguistics, 28*(4), 661-677. doi: 10.1017/S014271640707035X

Norris, J. H., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning, 50*(3), 417-528. doi: 10.1111/0023-8333.00136

Oppenheimer, J., & Arnaz, D. (Producers). (1951). *I Love Lucy* [Television series]. New York: Columbia Broadcasting System.

Otwinowska, A. (2016). *Cognate vocabulary in language acquisition and use: Attitudes awareness, activation*. Bristol: Multilingual Matters.

Paas, F., Renkl, A., & Sweller, J. (2004). Cognitive load theory: Instructional implications of the interaction between information structures and cognitive architecture. *Instructional Science, 32*(1-2), 1-8. doi: 10.1023/B:TRUC.0000021806.17516.d0

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

Paivio, A. (2007). *Mind and its evolution: A dual coding theoretical approach*. Mahwah: Lawrence Erlbaum Associates.

Paivio, A. (2010). Dual coding theory and the mental lexicon. *The Mental Lexicon, 5*(2), 205-230. doi: 10.1075/ml.5.2.04pai

Pallant, J. (2013). *SPSS survival manual. A step by step guide to data analysis using IBM SPSS* (fifth edition). Maidenhead: McGraw-Hill.

Park, M. (2004). *The effects of partial captions on Korean EFL learners' listening comprehension*. (Doctoral dissertation, University of Texas at Austin, Austin, United States). Retrieved from: https://repositories.lib.utexas.edu/handle/2152/12695

Parks, C. (1994). *Closed captioned TV: A resource for ESL literacy education*. Washington, DC: ERIC Systems Publications (no. ED372662). Retrieved from: https://files.eric.ed.gov/fulltext/ED372662.pdf

Pellicer-Sánchez, A., & Schmitt, N. (2010). Incidental vocabulary acquisition from an authentic novel: Do *Things Fall Apart*? *Reading in a Foreign Language, 22*(1), 31-55. Retrieved from: http://nflrc.lll.hawaii.edu/rfl/April2010/articles/pellicersanchez.pdf

Penno, J. F., Wilkinson, I. A. G., & Moore, D. W. (2002). Vocabulary acquisition from teacher explanation and repeated listening to stories: Do they overcome the Matthew effect? *Journal of Educational Psychology, 94*(1), 23-33. doi: 10.1037/0022-0663.94.1.23

Peters, E., Heynen, E., & Puimège, E. (2016). Learning vocabulary through audiovisual input: The differential effect of L1 subtitles and captions. *System, 63*, 134-148. doi: 10.1016/j.system.2016.10.002

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. doi: 10.1017/S0272263117000407

Pigada, M., & Schmitt, N. (2006). Vocabulary acquisition from extensive reading: A case study. *Reading in a Foreign Language, 18*(1), 1-28. Retrieved from: http://nflrc.hawaii.edu/rfl/April2006/pigada/pigada.html

Plass, J. L., Chun, D. M., Mayer, R. E., & Leutner, D. (1998). Supporting visual and verbal learning preferences in a second-language multimedia learning environment. *Journal of Educational Psychology, 90*(1), 25-36. doi: 10.1037/0022-0663.90.1.25

Plonsky, L. (2015, June). *Sample demographics and generalizability in SLA: A systematic review*. Paper presented at the Workshop on Reliability and Validity in SLA Research, Barcelona, Spain.

Price, K. (1983). Closed-captioned TV: An untapped resource. *Newsletter - Massachusetts Association of Teachers of English to Speakers of Other Languages, 12*(2), 1-8.

Pujadas, G. (forthcoming). *Learning language through TV series. Vocabulary, viewing comprehension, imagery and learners' perceptions* (Doctoral dissertation, Universitat de Barcelona, Barcelona, Spain).

Pujadas, G., & Muñoz, C. (2017, April). *Learning through subtitles. Learners' preferences and task perception*. Paper presented at the Seventh International Conference on Task-Based Language Teaching (TBLT), Barcelona, Spain.

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. doi: 10.1080/09571736.2019.1616806

Pulido, D. (2000). *The impact of topic familiarity, L2 reading proficiency, and L2 passage sight vocabulary on incidental vocabulary gain through reading for adult learners of Spanish as a foreign language*. (Doctoral dissertation, University of Illinois at Urbana-Champaign, Urbana-Champaign, United States). Retrieved from: https://www.ideals.illinois.edu/handle/2142/86189

Pulido, D. (2003). Modeling the role of second language proficiency and topic familiarity in second language incidental vocabulary acquisition through reading. *Language Learning, 53*(2), 233–284. doi: 10.1111/1467-9922.00217

Pulido, D. (2004). The relationship between text comprehension and second language incidental vocabulary acquisition: A matter of topic familiarity? *Language Learning, 54*(3), 469-523. doi: 10.1111/j.0023-8333.2004.00263.x

Qian, D. D. (2002). Investigating the relationship between vocabulary knowledge and academic reading performance: An assessment perspective. *Language Learning, 52*(3), 513-536. doi: 10.1111/1467-9922.00193

Quinn, E., Nation, I. S. P., & Millett, S. (2007). *Asian and Pacific speed readings for ESL learners. Twenty passages written at the one thousand word level*. Wellington: School of Linguistics and Applied Language Studies, Victoria University of Wellington.

Raine, P. (2012). *Incidental learning of vocabulary through subtitled authentic videos*. (Master's thesis, University of Birmingham, Birmingham, United Kingdom). Retrieved from: https://www.birmingham.ac.uk/index.aspx

Rasekh, Z. E., & Ranjbary, R. (2003). Metacognitive strategy training for vocabulary learning. *The Electronic Journal for English as a Second Language, 7*(2), 1-15. Retrieved from: http://tesl-ej.org/ej26/a5.html

Rice, M. L., Huston, A. C., Truglio, R., & Wright, J. (1990). Words from "Sesame Street": Learning vocabulary while viewing. *Developmental Psychology, 26*(3), 421-428. doi: 10.1037/0012-1649.26.3.421

Rieder, A. (2003). Implicit and explicit learning in incidental vocabulary acquisition. *Views, 12*(2), 24-39. Retrieved from: https://www.univie.ac.at/Anglistik/views/03_2/RIE_SGLE.PDF

Rodgers, M. P. H. (2013). *English language learning through viewing television: An investigation of comprehension, incidental vocabulary acquisition, lexical coverage, attitudes, and captions* (Doctoral dissertation, Victoria University of Wellington, Wellington, New Zealand). Retrieved from: http://researcharchive.vuw.ac.nz/

Rodgers, M. P. H. (2018). The images in television programs and the potential for learning unknown words. The relationship between on-screen imagery and vocabulary. *ITL - International Journal of Applied Linguistics, 169*(1), 191-211. doi: 10.1075/itl.00012.rod

Rodgers, M.P.H. Personal interview. 25 March 2018.

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

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

Rodriguez, M. C. (2005). Three options are optimal for multiple-choice items: A meta-analysis of 80 years of research. *Educational Measurement: Issues and Practice, 24*(2), 3-13. doi: 10.1111/j.1745-3992.2005.00006.x

Rott, S. (1999). The effect of exposure frequency on intermediate language learners' incidental vocabulary acquisition and retention through reading. *Studies in Second Language Acquisition, 21*(4), 589-619. doi: 10.1017/S0272263199004039

Sadoski, M. (2015). Dual Coding Theory. In J. M. Spector (Ed.), *The SAGE encyclopedia of educational technology* (pp. 238-239). Thousand Oaks: SAGE Publications.

Schmitt, N. (2000). *Vocabulary in language teaching*. Cambridge: Cambridge University Press.

Schmitt, N. (2008). Review article: Instructed second language vocabulary learning. *Language Teaching Research, 12*(3), 329-363. doi: 10.1177/1362168808089921

Schmitt, N., & Schmitt, D. (1995). Vocabulary notebooks: theoretical underpinnings and practical suggestions. *ELT Journal, 49*(2), 133-143. doi: 10.1093/elt/49.2.133

Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behavior of two new versions of the Vocabulary Levels Test. *Language Testing, 18*(1), 55-88. doi: 10.1177/026553220101800103

Schroeders, U., Wilhelm, O., & Bucholtz, N. (2010). Reading, listening and viewing comprehension in English as a foreign language: One or more constructs? *Intelligence, 38*(6), 562-573. doi: 10.1016/j.intell.2010.09.003

Shefelbine, J. L. (1990). Student factors related to variability in learning word meanings from context. *Journal of Literary Research, 22*(1), 71-97. doi: 10.1080/10862969009547695

Simons, G. F., & Fennig, C. D. (Eds.). (2018). *Ethnologue: Languages of Africa and Europe* (twenty-first edition). Dallas: SIL International Publications.

Singleton, D. (1999). *Exploring the second language mental lexicon*. Cambridge: Cambridge University Press.

Spada, N. (2005). Conditions and challenges in developing school-based SLA research programs. *The Modern Language Journal, 89*(3), 328-338. doi: 10.1111/j.1540-4781.2005.00308.x

Springfield! Springfield!. (n.d.). Springfield! Springfield!. [online] Retrieved from: https://www.springfieldspringfield.co.uk/ [Last accessed 24 June 2019].

Staehr, L. S. (2009). Vocabulary knowledge and advanced listening comprehension in English as a foreign language. *Studies in Second Language Acquisition, 31*(4), 577-607. doi: 10.1017/S0272263109990039

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly, 21*(4), 360-407. Retrieved from: https://rsrc.psychologytoday.com/files/u81/Stanovich__1986_.pdf

Stewart, M. A., & Pertusa, I. (2004). Gains to language learners from viewing target language closed-captioned films. *Foreign Language Annals, 37*(3), 438-442. doi: 10.1111/j.1944-9720.2004.tb02701.x

Suárez, M. M., & Gesa, F. (2019). Learning vocabulary with the support of sustained exposure to captioned video: do proficiency and aptitude make a difference? *The Language Learning Journal, 47*(4), 497-517. doi: 10.1080/09571736.2019.1617768

Swain, M. (1993). The Output Hypothesis: Just speaking and writing aren't enough. *The Canadian Modern Language Review / La Revue canadienne des langues vivantes, 50*(1), 158-164. doi: 10.3138/cmlr.50.1.158

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*, 257-285. doi: 10.1207/s15516709cog1202_4

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction, 4*(4), 295-312. doi: 10.1016/0959-4752(94)90003-5

Sweller, J. (1999). *Instructional design in technical areas*. Camberwell: Australian Council for Educational Research Press.

Swenson, E., & West, M. P. (1934). *On the counting of new words in textbooks for teaching foreign languages*. Toronto: The University of Toronto Press.

Sydorenko, T. (2010). Modality of input and vocabulary acquisition. *Language Learning & Technology 14*(2), 50-73. doi: 0125/44214

Talaván, N. (2007). Learning vocabulary through authentic video and subtitles. *TESOL-Spain Newsletter, 31*, 5-8.

Talaván, N. (2012). Justificación teórico-práctica del uso de los subtítulos en la enseñanza-aprendizaje de lenguas extranjeras. *TRANS. Revista de Traductología, 16*, 23-37. Retrieved from: http://www.trans.uma.es/trans_16/Trans16_023-037.pdf

Thompson, I., & Rubin, J. (1996). Can strategy instruction improve listening comprehension? *Foreign Language Annals, 29*(3), 331-342. doi: 10.1111/j.1944-9720.1996.tb01246.x

Tight, D. G. (2010). Perceptual learning style matching L2 vocabulary acquisition. *Language Learning, 60*(4), 792-833. doi: 10.1111/j.1467-9922.2010.00572.x

Tonzar, C., Lotto, L., & Job, R. (2009). L2 vocabulary acquisition in children: Effects of learning method and cognate status. *Language Learning, 59*(3), 623-646. doi: 10.1111/j.1467-9922.2009.00519.x

Tragant, E., Muñoz, C., & Spada, N. (2016). Maximizing young learners' input: An intervention program. *The Canadian Modern Language Review / La Revue canadienne des langues vivantes, 72*(2), 234-257. doi: 10.3138/cmlr.2942

Tschirner, E. (2001). Language acquisition in the classroom: The role of digital video. *Computer Assisted Language Learning, 14*(3-4), 305-319. doi: 10.1076/call.14.3.305.5796

Tse, C.-S., & Altarriba, J. (2014). The relationship between language proficiency and attentional control in Cantonese-English bilingual children: evidence from Simon, Simon switching, and working memory tasks. *Frontiers in Psychology, 5*(954), 1-18. doi: 10.3389/fpsyg.2014.00954

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. doi: 10.1111/lang.12343

Vandergrift, L. (2004). Listening to learn or learning to listen? *Annual Review of Applied Linguistics, 24*, 3-25. doi: 10.1017/S0267190504000017

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

Vanderplank, R. (1990). Paying attention to the words: Practical and theoretical problems in watching television programmes with uni-lingual (CEEFAX) sub-titles. *System, 18*(2), 221-234. doi: 10.1016/0346-251X(90)90056-B

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

Vanderplank, R. (2015). Thirty years of research into captions / same language subtitles and second / foreign language learning: Distinguishing between 'effects of' subtitles and 'effects with' subtitles for future research. In Y. Gambier, A. Caimi, & C. Mariotti (Eds.), *Subtitles and language learning. Principles, strategies and practical experiences* (pp. 19-40). Bern: Peter Lang.

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

Van der Slik, F. W. P. (2010). Acquisition of Dutch as a second language: The explanative power of cognate and genetic linguistic distance measures for 11 West European first languages. *Studies in Second Language Acquisition, 32*(3), 401-432. doi: 10.1017/S0272263110000021

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. doi: 10.1348/000709905X38946

van Merriënboer, J. J. G., & Ayres, P. (2005). Research on cognitive load theory and its design implications for e-learning. *Educational Technology Research and Development, 53*(3), 5-13. doi: 10.1007/BF02504793

van Merriënboer, J. J. G., Schuurman, J. G., de Croock, M. B. M., & Paas, F. G. W. C. (2002). Redirecting learners' attention during training: Effects on cognitive load, transfer test performance and training efficiency. *Learning and Instruction, 12*(1), 11–37. doi: 10.1016/S0959-4752(01)00020-2

van Zeeland, H., & Schmitt, N. (2013a). Incidental vocabulary acquisition through L2 listening: A dimensions approach. *System, 41*, 609-624. doi: 10.1016/j.system.2013.07.012

van Zeeland, H., & Schmitt, N. (2013b). Lexical coverage in L1 and L2 listening comprehension: The same or different from reading comprehension? *Applied Linguistics, 34*(4), 457-479. doi: 10.1093/applin/ams074

Verdú, V. (1996). *El planeta americano*. Barcelona: Anagrama.

Vidal, K. (2003). Academic listening: A source of vocabulary acquisition? *Applied Linguistics, 24*(1), 56-89. doi: 10.1093/applin/24.1.56

Vidal, K. (2011). A comparison of the effects of reading and listening on incidental vocabulary acquisition. *Language Learning, 61*(1), 219-258. doi: 10.1111/j.1467-9922.2010.00593.x

Wang, Y-C. (2012). Learning L2 vocabulary with American TV drama. From the learners' perspective. *English Language Teaching, 5*(8), 217-225. doi: 10.5539/elt.v5n8p217

Waring, R., & Takaki, M. (2003). At what rate do learners learn and retain new vocabulary from reading a graded reader? *Reading in a Foreign Language, 15*(2), 130-163. Retrieved from: http://nflrc.hawaii.edu/rfl/October2003/waring/waring.html

Webb, S. (2007). The effects of repetition on vocabulary knowledge. *Applied Linguistics, 28*(1), 46-65. doi:10.1093/applin/aml048

Webb, S. (2008). Receptive and productive vocabulary sizes of L2 learners. *Studies in Second Language Acquisition, 30*(1), 79-95. doi: 10.1017/S0272263108080042

Webb, S. (2011). Selecting television programs for language learning: Investigating television programs from the same genre. *International Journal of English Studies, 11*(1), 117-135. doi: 10.6018/ijes/2011/1/137131

Webb, S. (2015). Extensive viewing: Language learning through watching television. In D. Nunan, & J. C. Richards (Eds.), *Language learning beyond the classroom* (pp. 159-168). New York: Routledge.

Webb, S., & Chang, A. C.-S. (2015a). How does prior word knowledge affect vocabulary learning progress in an extensive reading program? *Studies in Second Language Acquisition*, *37*(4), 651-675. doi: 10.1017/S0272263114000606

Webb, S., & Chang, A. C.-S. (2015b). Second language vocabulary learning through extensive reading with audio support: How do frequency and distribution of occurrence affect learning? *Language Teaching Research, 19*(6), 667-686. doi: 10.1177/1362168814559800

Webb, S., & Nation, P. (2017). *How vocabulary is learned*. Oxford: Oxford University Press.

Webb, S., & Rodgers, M. P. H. (2009a). The lexical coverage of movies. *Applied Linguistics, 30*(3), 407-427. doi: 10.1093/applin/amp010

Webb, S., & Rodgers, M. P. H. (2009b). Vocabulary demands of television programs. *Language Learning, 59*(2), 335-366. doi: 10.1111/j.1467-9922.2009.00509.x

Wesche, M., & Paribakht, T. S. (1996). Assessing second language vocabulary knowledge: Depth versus breadth. *The Canadian Modern Language Review / La Revue canadienne des langues vivantes, 53*(1), 13-40. doi: 10.3138/cmlr.53.1.13

Williams, H., & Thorne, D. (2000). The value of teletext subtitling as a medium for language learning. *System, 28*(2), 217-228. doi: 10.1016/S0346-251X(00)00008-7

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. doi: 10125/44203

Winke, P., Gass, S., & Sydorenko, T. (2013). Factors influencing the use of captions by foreign language learners: An eye-tracking study. *The Modern Language Journal, 97*(1), 254-275. doi: 10.1111/j.1540-4781.2012.01432.x

Yoshii, M., & Flaitz, J. (2002). Second language incidental vocabulary retention: The effect of text and picture annotation types. *CALICO Journal, 20*(1), 33-58. doi: 10.1558/cj.v20i1.33-58

Yuksel, D., & Tanriverdi, B. (2009). Effects of watching captioned movie clip on vocabulary development of EFL learners. *The Turkish Online Journal of Educational Technology, 8*(2), 48-54. Retrived from: http://tojet.net/articles/v8i2/824.pdf

Zarei, A. A. (2009). The effect of bimodal, standard, and reversed subtitling on L2 vocabulary recognition and recall. *Pazhuhesh-e Zabanha-ye Khareji, 49*, 65-85. Retrieved from: http://www.ikiu.ac.ir/public-files/profiles/items/1310286471.pdf

Zarei, A. A., & Rashvand, Z. (2011). The effect of interlingual and intralingual verbatim and nonverbatim subtitles on L2 vocabulary comprehension and production. *Journal of Language Teaching and Research, 2*(3), 618-625. doi:10.4304/jltr.2.3.618-625

Zhang, X. (2013). The I don't know option in the vocabulary size test. *TESOL Quarterly, 47*(4), 790-811. doi: 10.1002/tesq.98

APPENDIX A - BACKGROUND QUESTIONNAIRE

A.1. - Grade 6

Qüestionari: Aprenentatge de l'anglès

Apreciats/des pares, mares i famílies,

Els estaríem molt agraïts si ens poguessin ajudar tot contestant les següents preguntes sobre l'aprenentatge de l'anglès dels seus fills/es. El Grup de Recerca GRAL de la Universitat de Barcelona (dins d'un projecte de recerca finançat pel Ministerio) porta a terme aquest estudi per tal d'entendre millor el contacte que tenen els/les nens/es amb l'anglès fora de l'aula. Els resultats ens ajudaran a entendre com aprenen anglès i com se'ls pot ajudar a aprendre aquest idioma més ràpidament. El qüestionari consta de 5 parts i té una durada aproximada de 15 minuts. Tota la informació proporcionada es tractarà de forma CONFIDENCIAL i només s'utilitzarà amb finalitats de recerca.

Moltes gràcies per la seva participació!

A. Ús de l'anglès fora de l'aula

1. Indiqui la freqüència amb la qual el seu fill/a realitza les següents activitats.

1a. Veure pel·lícules, dibuixos i/o sèries en anglès.

	Mai	Menys d'1 cop / mes	Entre 1-3 cops / mes	Entre 1-3 cops / setmana	Entre 4-6 cops / setmana	Cada dia
Sense subtítols						
Amb subtítols en català / castellà						
Amb subtítols en anglès						

1b. Jugar a jocs d'ordinador, mòbil o videojocs en anglès.

	Mai	Menys d'1 cop / mes	Entre 1-3 cops / mes	Entre 1-3 cops / setmana	Entre 4-6 cops / setmana	Cada dia
Sense subtítols						
Amb subtítols en català / castellà						
Amb subtítols en anglès						

2. Indiqui la freqüència amb la qual el seu fill/a realitza les següents activitats.

	Mai	Menys d'1 cop / mes	Entre 1-3 cops / mes	Entre 1-3 cops / setmana	Entre 4-6 cops / setmana	Cada dia
Escoltar música en anglès (ex: ràdio, CDs, iPhone, etc.)						
Llegir llibres, revistes o còmics en anglès						

3. Quan i amb quina freqüència el seu fill/a parla en anglès cara a cara?

	Mai	Menys d'1 cop / mes	Entre 1-3 cops / mes	Entre 1-3 cops / setmana	Entre 4-6 cops / setmana	Cada dia
Amb amics						
Amb familiars						
Amb turistes						
A l'estranger						

4. Indiqui la freqüència amb la qual el seu fill/a realitza les següents activitats per Internet.

	Mai	Menys d'1 cop / mes	Entre 1-3 cops / mes	Entre 1-3 cops / setmana	Entre 4-6 cops / setmana	Cada dia
Parlar en anglès amb algú (ex: Skype)						
Escriure en suport digital en anglès (ex: emails, xats, WhatsApp, Facebook...)						
Llegir textos en anglès (ex: revistes, pàgines web, blocs...)						
Mirar vídeos en anglès a YouTube						
Escoltar la ràdio / podcasts / música en anglès						

5. Si el seu fill/a realitza altres activitats per tal d'aprendre o practicar l'anglès, si us plau completi la següent frase.

Per tal d'aprendre o practicar l'anglès, el meu fill/a...

____

B. Classes extraescolars

6. El seu fill/a assisteix o ha assistit a classes extraescolars d'anglès (dins o fora del centre escolar)?

___ Sí (passi a la pregunta 6a) ___ No (passi a la pregunta 7)

6a. Nombre d'anys que el seu fill/a ha assistit a classes extraescolars d'anglès.

__________ anys

6b. Indiqui la freqüència (en hores/setmana) de les classes extraescolars d'anglès a les quals assisteix / ha assistit el seu fill/a.

___ 1h / setmana ___ 1,5h / setmana ___ 2h / setmana

___ 3h / setmana ___ Més de 3h / setmana

___ Altres: ____

6c. A quina edat el seu fill/a va començar a anar a classes extraescolars d'anglès?

___ 0-2 anys ___ 3-4 anys ___ 5-6 anys

___ 7-8 anys ___ 9-10 anys ___ 11-12 anys

6d. Mencioni la raó (o raons) per la qual / les quals van apuntar el seu fill/a a classes extraescolars d'anglès. (Pot marcar més d'una opció)

___ Perquè li agrada l'anglès.

___ Perquè li faltava més pràctica oral.

___ Perquè va suspendre l'assignatura curricular d'anglès.

___ Perquè el/la professor/a de l'escola ens ho va recomanar.

___ Altres: ____

C. Estades a l'estranger

7. El seu fill/a ha estat alguna vegada a l'estranger per un període superior a dues setmanes durant el qual va utilitzar l'anglès habitualment?

___ Sí (passi a la pregunta 7a) ___ No (passi a la pregunta 8)

7a. Indiqui quantes vegades el seu fill/a ha estat a l'estranger per un període superior a dues setmanes durant el qual va utilitzar l'anglès habitualment.

___ 1 vegada ___ 2 vegades ___ 3 vegades

___ Altres: ____

7b. Si us plau, especifiqui la durada de cada estada a l'estranger. (Per exemple, 1a vegada = 15 dies; 2a vegada = 20 dies; etc.)

____

7c. En general, durant les estades a l'estranger, el seu fill/a...

	Menys del 25% del temps	Entre més del 25% i el 50% del temps	Entre més del 50% i el 75% del temps	Més del 75% del temps
Va escoltar parlar en anglès				
Va parlar en anglès				

D. Campaments de llengües

8. El seu fill/a ha estat alguna vegada en campaments on s'utilitzava l'anglès habitualment?

___ Sí (passi a la pregunta 8a) ___ No (passi a la pregunta 9)

8a. Indiqui quantes vegades el seu fill/a ha estat en campaments de llengües.

___ 1 vegada ___ 2 vegades ___ 3 vegades

___ Altres: ____

8b. Si us plau, especifiqui la durada de cada campament. (Per exemple, 1a vegada = 7 dies; 2a vegada = 10 dies; etc.)

____

8c. En general, durant els campaments de llengües on s'utilitzava l'anglès habitualment, el seu fill/a...

	Menys del 25% del temps	Entre més del 25% i el 50% del temps	Entre més del 50% i el 75% del temps	Més del 75% del temps
Va escoltar parlar en anglès				
Va parlar en anglès				

E. Entorn familiar

Si cap de les opcions que es donen és vàlida, seleccioni ‘NO PROCEDEIX’.

9. Nivell d’estudis completats

9a. Mare

___ Educació Primària ___ Educació Secundària

___ Cicle formatiu de grau mig i/o superior ___ Estudis universitaris

___ No procedeix

___ Altres: ____

9b. Pare

___ Educació Primària ___ Educació Secundària

___ Cicle formatiu de grau mig i/o superior ___ Estudis universitaris

___ No procedeix

___ Altres: ____

9c. Altres persones amb les quals el seu fill/a convisqui (ex: avis/àvies, parelles)

___ Educació Primària ___ Educació Secundària

___ Cicle formatiu de grau mig i/o superior ___ Estudis universitaris

___ No procedeix

___ Altres: ____

10. Quina és la seva feina actual?

10a. Mare (si no procedeix, escrigui ‘NO’). Si actualment es troba a l’atur, indiqui també la feina que tenia abans.

____

10b. Pare (si no procedeix, escrigui ‘NO’). Si actualment es troba a l’atur, indiqui també la feina que tenia abans.

____

10c. Altres persones amb les quals el seu fill/a convisqui (si no procedeix, escrigui ‘NO’). Si actualment es troba a l’atur, indiqui també la feina que tenia abans.

____

11. Té la necessitat d'usar l'anglès en el seu lloc de treball?

11a. Mare

___ Sí ___ No ___ No procedeix

11b. Pare

___ Sí ___ No ___ No procedeix

11c. Altres persones amb les quals el seu fill/a convisqui (ex: avis/àvies, parelles)

___ Sí ___ No ___ No procedeix

12. Es sent còmode/a usant l'anglès (parlant, escrivint, etc.)?

12a. Mare

___ Sí ___ No ___ No procedeix

12b. Pare

___ Sí ___ No ___ No procedeix

12c. Altres persones amb les quals el seu fill/a convisqui (ex: avis/àvies, parelles)

___ Sí ___ No ___ No procedeix

13. Quina llengua o llengües es parlen a casa?

___ Català ___ Castellà ___ Català i castellà

___ Altres: ____

14. Edat del seu fill/a: __________ anys

15. Sexe del seu fill/a: ___ Home ___ Dona

16. Nom del centre: ____

17. Especifiqui la nota d'anglès que el seu fill/a va obtenir en la darrera avaluació.

___ Suspens (0-4,9) ___ Aprovat (5-6,9)

___ Notable (7-8,9) ___ Excel·lent (9-10)

18. Comentaris: Si us plau, utilitzi aquest espai per a fer algun comentari sobre qualsevol aspecte relacionat amb el contacte amb l'anglès que el seu fill/a té fora de l'aula que no s'hagi tractat en les preguntes anteriors.

____

____

____

MOLTES GRÀCIES PER PARTICIPAR!

A.2. - Grade 10 and university

Qüestionari: Aprenentatge de l'anglès

Benvolguts/des estudiants,

Us estaríem molt agraïts si ens poguéssiu ajudar tot contestant les següents preguntes sobre l'aprenentatge de l'anglès. El Grup de Recerca GRAL de la Universitat de Barcelona (dins d'un projecte de recerca finançat pel Ministerio) porta a terme aquest estudi per tal d'entendre millor el rol de l'exposició a l'anglès –o contacte amb l'anglès– fora de l'aula. El qüestionari consta de 5 parts i té una durada aproximada de 15 minuts. Tota la informació proporcionada es tractarà de forma CONFIDENCIAL i només s'utilitzarà amb finalitats de recerca.

Moltes gràcies per la vostra participació!

A. Ús de l'anglès fora de l'aula

1. Indica la freqüència amb la qual realitzes les següents activitats.

1a. Veure pel·lícules i sèries en anglès en versió original.

	Mai	Menys d'1 cop / mes	Entre 1-3 cops / mes	Entre 1-3 cops / setmana	Entre 4-6 cops / setmana	Cada dia
Amb subtítols en català / castellà						
Amb subtítols en anglès						
Sense subtítols						

1b. Jugar a videojocs en anglès.

	Mai	Menys d'1 cop / mes	Entre 1-3 cops / mes	Entre 1-3 cops / setmana	Entre 4-6 cops / setmana	Cada dia
Individual						
Multijugador						
MMO (multijugador massiu)						

1c. Indica la freqüència amb la qual realitzes les següents activitats.

	Mai	Menys d'1 cop / mes	Entre 1-3 cops / mes	Entre 1-3 cops / setmana	Entre 4-6 cops / setmana	Cada dia
Escoltar música en anglès (ex: ràdio, CDs, iPhone, etc.)						
Llegir llibres, revistes o còmics en anglès						

1d. Quan i amb quina freqüència parles en anglès cara a cara?

	Mai	Menys d'1 cop / mes	Entre 1-3 cops / mes	Entre 1-3 cops / setmana	Entre 4-6 cops / setmana	Cada dia
Amb amics						
Amb familiars						
Amb turistes						
A l'estranger						

1e. Indica la freqüència amb la qual realitzes les següents activitats per Internet.

	Mai	Menys d'1 cop / mes	Entre 1-3 cops / mes	Entre 1-3 cops / setmana	Entre 4-6 cops / setmana	Cada dia
Parlar en anglès amb algú (ex: Skype)						
Escriure en suport digital (ex: emails, xats, WhatsApp, Facebook, Twitter)						
Llegir textos (ex: ebooks, revistes, pàgines web, blocs, diaris, manuals d'usuari)						
Mirar vídeos a YouTube						
Escoltar la ràdio / podcasts / música a Spotify						

2. Especifica si portes a terme les accions següents quan fas aquestes activitats. Si no realitzes alguna d'aquestes activitats regularment, selecciona "no procedeix".

2a. Quan escolto música en anglès...

	Mai	A vegades	Sovint	Molt sovint	Sempre	No procedeix
Intento entendre la cançó en general						
Intento segmentar les paraules						

2b. Quan veig pel·lícules en anglès amb subtítols en català / castellà...

	Mai	A vegades	Sovint	Molt sovint	Sempre	No procedeix
Intento escoltar l'àudio abans de llegir els subtítols						
Llegeixo els subtítols abans d'escoltar l'àudio						
Només llegeixo els subtítols si no entenc l'àudio						

2c. Quan veig pel·lícules en anglès amb subtítols en anglès...

	Mai	A vegades	Sovint	Molt sovint	Sempre	No procedeix
Intento escoltar l'àudio abans de llegir els subtítols						
Llegeixo els subtítols abans d'escoltar l'àudio						
Només llegeixo els subtítols si no entenc l'àudio						

2d. Quan navego per Internet...

	Mai	A vegades	Sovint	Molt sovint	Sempre	No procedeix
Utilitzo un traductor online per entendre un text complet en anglès						
Utilitzo un traductor online per entendre paraules individuals en anglès						

3. Si realitzes altres activitats per tal d'aprendre o practicar l'anglès quan veus pel·lícules i sèries en anglès, si us plau completa la següent frase.

Per tal d'aprendre o practicar l'anglès, quan veig pel·lícules, sèries, programes o vídeos en anglès, jo...

____

B. Estades a l'estranger

4. Has estat alguna vegada a l'estranger per un període superior a dues setmanes, durant el qual vas utilitzar l'anglès habitualment?

____ Sí (passa a la pregunta 4a) ____ No (passa a la pregunta 5)

4a. Si us plau, especifica la durada de la teva estada/es a l'estranger durant la qual / les quals vas utilitzar l'anglès habitualment. (Selecciona "no procedeix" si no tens una segona i/o tercera estada a l'estranger)

	Més de 2 setmanes i menys de 4 setmanes	Entre 1 mes i menys d'1,5 mesos	Entre 1,5 mesos i menys de 3 mesos	Entre 3 i 6 mesos	Més de 6 mesos	No procedeix
Estada més llarga						
2a estada més llarga						
3a estada més llarga						

4b. Durant la meva estada més llarga a l'estranger...

	Menys del 25% del temps	Entre més del 25% i el 50% del temps	Entre més del 50% i el 75% del temps	Més del 75% del temps
Vaig escoltar parlar en anglès				
Vaig parlar en anglès				

4c. Durant la meva 2a estada més llarga a l'estranger...

	Menys del 25% del temps	Entre més del 25% i el 50% del temps	Entre més del 50% i el 75% del temps	Més del 75% del temps
Vaig escoltar parlar en anglès				
Vaig parlar en anglès				

4d. Durant la meva 3a estada més llarga a l'estranger...

	Menys del 25% del temps	Entre més del 25% i el 50% del temps	Entre més del 50% i el 75% del temps	Més del 75% del temps
Vaig escoltar parlar en anglès				
Vaig parlar en anglès				

C. Campaments de llengües

5. Has estat alguna vegada en campaments on s'utilitzava l'anglès habitualment?

___ Sí (passa a la pregunta 5a) ___ No (passa a la pregunta 6)

5a. Indica quantes vegades has estat en campaments de llengües.

___ 1 vegada ___ 2 vegades ___ 3 vegades

___ Altres: ____

5b. Si us plau, especifica la durada de cada campament. (Per exemple, 1a vegada = 10 dies; 2a vegada = 15 dies; etc.)

____

5c. En general, durant els campaments de llengües on s'utilitzava l'anglès habitualment...

	Menys del 25% del temps	Entre més del 25% i el 50% del temps	Entre més del 50% i el 75% del temps	Més del 75% del temps
Vaig escoltar parlar en anglès				
Vaig parlar en anglès				

D. Classes extraescolars

6. Alguna vegada has anat a classes extraescolars d'anglès (dins o fora del centre escolar)?

___ Sí (passa a la pregunta 6a) ___ No (passa a la pregunta 7)

6a. Nombre d'anys que has assistit a classes extraescolars d'anglès.

__________ anys

6b. Indica quan vas anar a classes extraescolars d'anglès. (Pots seleccionar més d'una opció)

___ Educació Primària ___ Educació Secundària ___ Batxillerat

___ Cicle Formatiu ___ Universitat

6c. En general, indica la freqüència (en hores/setmana) de les classes extraescolars d'anglès.

___ 1h / setmana ___ 1,5h / setmana ___ 2h / setmana

___ 3h / setmana ___ Més de 3h / setmana

___ Altres: ____

6d. Menciona la raó (o raons) per la qual / les quals vas anar a classes extraescolars d'anglès. (Pots marcar més d'una opció)

___ Perquè m'agrada l'anglès.

___ Perquè em feia falta més pràctica oral.

___ Perquè vaig suspendre l'assignatura curricular d'anglès.

___ Altres: ____

E. Informació general

7. Edat: __________ anys

8. Sexe: ___ Home ___ Dona

9. Especifiqui la nota obtinguda en l'últim curs d'anglès que vas fer.

___ Suspens (0-4,9) ___ Aprovat (5-6,9)

___ Notable (7-8,9) ___ Excel·lent (9-10)

10. Què estàs estudiant ara? (Ex.: educació secundària, batxillerat en ciències, grau en estudis anglesos, etc.)

____

11. A quin curs estàs? (Ex.: 1r, 2n, etc.)

____

12. Nom del centre: ____

13. Comentaris: Si us plau, utilitza aquest espai per a comentar sobre qualsevol aspecte relacionat amb el teu contacte amb l'anglès fora de l'aula que no s'hagi tractat en les preguntes anteriors. [Ex.: tenir familiars de parla anglesa (pare, mare, cosins/es, etc.)]

____

____

____

MOLTES GRÀCIES PER PARTICIPAR!

APPENDIX B - VOCABULARY DEMANDS OF TV SERIES IN GRADE 6

Tables B.1-B.24 show the vocabulary demands of the 24 episodes selected in Grade 6. They indicate the raw number and the percentage of tokens, types and WFs found in each frequency band as well as the cumulative coverage in percentage.

Table B.1
Tokens, types, WFs and cumulative coverage of episode G6E01 ("Hotel Hangout").

Frequency band	Tokens		Types		WFs		Cumulative coverage (in %)
	Raw	%	Raw	%	Raw	%	
1k	2,210	90.80	478	73.88	375	72.25	90.80
2k	101	4.15	70	10.82	68	13.10	94.95
3k*	26	1.07	21	3.25	21	4.05	96.02
4k	22	0.90	18	2.78	15	2.89	96.92
5k	21	0.86	14	2.16	12	2.31	97.78
6k	5	0.21	5	0.77	5	0.96	97.99
7k*	6	0.25	4	0.62	4	0.77	98.24
8k	4	0.16	4	0.62	4	0.77	98.40
9k	3	0.12	3	0.46	3	0.58	98.52
10k	1	0.04	1	0.15	1	0.19	98.56
11-25k	15	0.61	12	1.84	11	2.11	99.17
Off-list	20	0.82	16	2.47	???		99.99
Total	2,434	100	647	100	≈ 519	???	≈ 100

Note. * Indicate the levels when 95% and 98% coverage were reached.

Table B.2
Tokens, types, WFs and cumulative coverage of episode G6E02 ("The Fairest of Them All").

Frequency band	Tokens		Types		WFs		Cumulative coverage (in %)
	Raw	%	Raw	%	Raw	%	
1k	2,121	90.29	474	73.04	383	71.86	90.29
2k	97	4.13	69	10.63	64	12.01	94.42
3k*	27	1.15	24	3.70	22	4.13	95.57
4k	20	0.85	16	2.47	15	2.81	96.42
5k	20	0.85	17	2.62	15	2.81	97.27
6k	10	0.43	8	1.23	8	1.50	97.70
7k*	12	0.51	5	0.77	5	0.94	98.21
8k	6	0.26	4	0.62	4	0.75	98.47
9k	4	0.17	3	0.46	3	0.56	98.64
10k	4	0.17	4	0.62	4	0.75	98.81
11-25k	12	0.51	10	1.51	10	1.89	99.32
Off-list	16	0.68	14	2.16	???		100
Total	2,349	100	649	100	≈ 533	???	≈ 100

Note. * Indicate the levels when 95% and 98% coverage were reached.

Table B.3
Tokens, types, WFs and cumulative coverage of episode G6E03 ("Maddie Checks In").

Frequency band	Tokens		Types		WFs		Cumulative coverage (in %)
	Raw	%	Raw	%	Raw	%	
1k	2,261	91.69	504	76.60	393	74.57	91.69
2k*	86	3.49	71	10.79	67	12.71	95.18
3k	21	0.85	17	2.58	17	3.23	96.03
4k	36	1.46	17	2.58	17	3.23	97.49
5k	9	0.36	7	1.06	6	1.14	97.85
6k*	20	0.81	12	1.82	11	2.09	98.66
7k	1	0.04	1	0.15	1	0.19	98.70
8k	3	0.12	3	0.46	2	0.38	98.82
9k	2	0.08	2	0.30	2	0.38	98.90
10k	2	0.08	2	0.30	2	0.38	98.98
11-25k	13	0.52	9	1.36	9	1.71	99.50
Off-list	12	0.49	12	1.82	???		99.99
Total	2,466	100	658	100	≈ 527	???	≈ 100

Note. * Indicate the levels when 95% and 98% coverage were reached.

Table B.4
Tokens, types, WFs and cumulative coverage of episode G6E04 ("Hotel Inspector").

Frequency band	Tokens		Types		WFs		Cumulative coverage (in %)
	Raw	%	Raw	%	Raw	%	
1k	2,235	88,66	496	70,16	377	67.56	88.66
2k	128	5.08	80	11.32	75	13.44	93.74
3k*	39	1.55	23	3.25	23	4.12	95.29
4k	29	1.15	23	3.25	22	3.94	96.44
5k	27	1.07	24	3.39	23	4.12	97.51
6k	10	0.40	10	1.41	9	1.61	97.91
7k	2	0.08	2	0.28	2	0.36	97.99
8k*	3	0.12	3	0.42	3	0.54	98.11
9k	2	0.08	2	0.28	2	0.36	98.19
10k	5	0.20	4	0.57	4	0.72	98.39
11-25k	18	0.72	18	2.52	18	3.24	99.11
Off-list	23	0.91	21	2.97	???		100
Total	2,521	100	707	100	≈ 558	???	≈ 100

Note. * Indicate the levels when 95% and 98% coverage were reached.

Table B.5
Tokens, types, WFs and cumulative coverage of episode G6E05 ("Grounded on the 23rd Floor").

Frequency band	Tokens		Types		WFs		Cumulative coverage (in %)
	Raw	%	Raw	%	Raw	%	
1k	2,140	91.34	481	74.81	375	72.96	91.34
2k	80	3.41	64	9.95	60	11.67	94.75
3k*	27	1.15	25	3.89	24	4.67	95.90
4k	19	0.81	17	2.64	17	3.31	96.71
5k	24	1.02	11	1.71	10	1.95	97.73
6k*	7	0.30	7	1.09	7	1.36	98.03
7k	5	0.21	4	0.62	4	0.78	98.24
8k	3	0.13	3	0.47	3	0.58	98.37
9k	3	0.13	2	0.31	2	0.39	98.50
10k	2	0.09	2	0.31	2	0.39	98.59
11-25k	11	0.46	10	1.57	10	1.93	99.05
Off-list	22	0.94	16	2.49	???		99.99
Total	2,343	100	643	100	≈ 514	???	≈ 100

Note. * Indicate the levels when 95% and 98% coverage were reached.

Table B.6
Tokens, types, WFs and cumulative coverage of episode G6E06 ("The Prince & The Plunger").

Frequency band	Tokens		Types		WFs		Cumulative coverage (in %)
	Raw	%	Raw	%	Raw	%	
1k	2,615	92.14	518	75.84	399	73.75	92.14
2k*	111	3.91	67	9.81	63	11.65	96.05
3k	23	0.81	20	2.93	20	3.70	96.86
4k	24	0.85	22	3.22	20	3.70	97.71
5k*	17	0.60	14	2.05	13	2.40	98.31
6k	6	0.21	5	0.73	5	0.92	98.52
7k	2	0.07	2	0.29	2	0.37	98.59
8k	6	0.21	4	0.59	4	0.74	98.80
9k	2	0.07	2	0.29	2	0.37	98.87
10k	-	-	-	-	-	-	98.87
11-25k	17	0.62	13	1.92	13	2.38	99.49
Off-list	15	0.53	15	2.20	???		100
Total	2,838	100	683	100	≈ 541	???	≈ 100

Note. * Indicate the levels when 95% and 98% coverage were reached.

Table B.7
Tokens, types, WFs and cumulative coverage of episode G6E07 ("Footloser").

Frequency band	Tokens		Types		WFs		Cumulative coverage (in %)
	Raw	%	Raw	%	Raw	%	
1k	2,427	92.67	488	76.37	373	74.30	92.67
2k*	73	2.79	58	9.08	56	11.16	95.46
3k	24	0.92	15	2.35	13	2.59	96.38
4k	24	0.92	19	2.97	18	3.59	97.30
5k	16	0.61	13	2.03	13	2.59	97.91
6k*	17	0.65	13	2.03	13	2.59	98.56
7k	2	0.08	2	0.31	2	0.40	98.64
8k	5	0.19	4	0.63	4	0.80	98.83
9k	1	0.04	1	0.16	1	0.20	98.87
10k	-	-	-	-	-	-	98.87
11-25k	9	0.35	9	1.41	9	1.80	99.22
Off-list	21	0.80	17	2.66	???		100
Total	2,619	100	639	100	≈ 502	???	≈ 100

Note. * Indicate the levels when 95% and 98% coverage were reached.

Table B.8
Tokens, types, WFs and cumulative coverage of episode G6E08 ("A Prom Story").

Frequency band	Tokens		Types		WFs		Cumulative coverage (in %)
	Raw	%	Raw	%	Raw	%	
1k	2,298	90.54	470	74.13	379	71.78	90.54
2k	77	3.03	59	9.31	59	11.17	93.57
3k	22	0.87	18	2.84	17	3.22	94.44
4k*	22	0.87	20	3.15	20	3.79	95.31
5k	29	1.14	18	2.84	17	3.22	96.45
6k	12	0.47	11	1.74	11	2.08	96.92
7k	10	0.39	6	0.95	6	1.14	97.31
8k	4	0.16	4	0.63	2	0.38	97.47
9k*	25	0.99	3	0.47	3	0.57	98.46
10k	6	0.24	4	0.63	4	0.76	98.70
11-25k	16	0.64	11	1.75	10	1.90	99.34
Off-list	17	0.67	9	1.42	???		100
Total	2,538	100	634	100	≈ 528	???	≈ 100

Note. * Indicate the levels when 95% and 98% coverage were reached.

Table B.9
Tokens, types, WFs and cumulative coverage of episode G6E09 ("Band in Boston").

Frequency band	Tokens		Types		WFs		Cumulative coverage (in %)
	Raw	%	Raw	%	Raw	%	
1k	2,500	91.11	474	76.45	367	75.05	91.11
2k*	119	4.34	58	9.35	54	11.04	95.45
3k	23	0.84	14	2.26	14	2.86	96.29
4k	21	0.77	15	2.42	11	2.25	97.06
5k	18	0.66	13	2.10	12	2.45	97.72
6k*	18	0.66	8	1.29	8	1.64	98.38
7k	8	0.29	5	0.81	5	1.02	98.67
8k	7	0.26	7	1.13	7	1.43	98.93
9k	3	0.11	3	0.48	3	0.61	99.04
10k	3	0.11	2	0.32	2	0.41	99.15
11-25k	8	0.29	6	0.96	6	1.22	99.44
Off-list	16	0.58	14	2.26	???		100
Total	2,744	100	620	100	≈ 489	???	≈ 100

Note. * Indicate the levels when 95% and 98% coverage were reached.

Table B.10
Tokens, types, WFs and cumulative coverage of episode G6E10 ("Cody Goes to Camp").

Frequency band	Tokens		Types		WFs		Cumulative coverage (in %)
	Raw	%	Raw	%	Raw	%	
1k	2,712	90.22	571	71.02	440	69.07	90.22
2k	119	3.96	90	11.19	85	13.34	94.18
3k*	38	1.26	30	3.73	29	4.55	95.44
4k	27	0.90	21	2.61	19	2.98	96.34
5k	19	0.63	19	2.36	18	2.98	96.97
6k	12	0.40	11	1.37	11	1.73	97.37
7k	9	0.30	7	0.87	7	1.10	97.67
8k	1	0.03	1	0.12	1	0.16	97.70
9k*	10	0.33	8	1	8	1.26	98.03
10k	4	0.13	4	0.50	4	0.63	98.16
11-25k	20	0.66	14	1.73	14	2.20	98.82
Off-list	35	1.16	27	3.36	???		99.98
Total	3,006	100	804	100	≈ 637	???	≈ 100

Note. * Indicate the levels when 95% and 98% coverage were reached.

Table B.11
Tokens, types, WFs and cumulative coverage of episode G6E11 ("To Catch a Thief").

Frequency band	Tokens		Types		WFs		Cumulative coverage (in %)
	Raw	%	Raw	%	Raw	%	
1k	2,608	91.35	557	76.41	426	74.35	91.35
2k*	108	3.78	72	9.88	67	11.69	95.13
3k	17	0.60	14	1.92	14	2.44	95.73
4k	31	1.09	19	2.61	18	3.14	96.82
5k*	36	1.26	18	2.47	16	2.79	98.08
6k	4	0.14	4	0.55	4	0.70	98.22
7k	3	0.11	3	0.41	3	0.52	98.33
8k	4	0.14	4	0.55	4	0.70	98.47
9k	7	0.25	4	0.55	4	0.70	98.72
10k	4	0.14	4	0.55	4	0.70	98.86
11-25k*	17	0.62	13	1.78	13	2.25	99.48
Off-list	16	0.56	16	2.19	???		100
Total	2,855	100	729	100	≈ 573	???	≈ 100

Note. * Indicate the levels when 95% and 98% coverage were reached.

Table B.12
Tokens, types, WFs and cumulative coverage of episode G6E12 ("It's a Mad, Mad, Mad Hotel").

Frequency band	Tokens		Types		WFs		Cumulative coverage (in %)
	Raw	%	Raw	%	Raw	%	
1k	2,101	89.79	476	74.26	365	71.29	89.79
2k	103	4.40	77	12.01	73	14.26	94.19
3k*	37	1.58	11	1.72	11	2.15	95.77
4k	34	1.45	26	4.06	23	4.49	97.22
5k*	22	0.94	13	2.03	13	2.54	98.16
6k	4	0.17	3	0.47	3	0.59	98.33
7k	3	0.13	3	0.47	3	0.59	98.46
8k	4	0.17	3	0.47	3	0.59	98.63
9k	3	0.13	2	0.31	2	0.39	98.76
10k	1	0.04	1	0.16	1	0.20	98.80
11-25k	17	0.73	15	2.35	15	3.15	99.53
Off-list	11	0.47	10	1.56	???		100
Total	2,340	100	641	100	≈ 512	???	≈ 100

Note. * Indicate the levels when 95% and 98% coverage were reached.

Table B.13
Tokens, types, WFs and cumulative coverage of episode G6E13 ("Poor Little Rich Girl").

Frequency band	Tokens		Types		WFs		Cumulative coverage (in %)
	Raw	%	Raw	%	Raw	%	
1k	2,521	92.04	546	76.58	422	75.36	92.04
2k*	91	3.32	63	8.84	59	10.54	95.36
3k	19	0.69	17	2.38	17	3.04	96.05
4k	25	0.91	19	2.66	18	3.21	96.96
5k	22	0.80	18	2.52	18	3.21	97.76
6k*	17	0.62	11	1.54	10	1.79	98.38
7k	4	0.15	3	0.42	3	0.54	98.53
8k	2	0.07	2	0.28	2	0.36	98.60
9k	-	-	-	-	-	-	98.78
10k	5	0.18	4	0.56	4	0.71	98.78
11-25k	7	0.26	7	0.98	7	1.26	99.04
Off-list	26	0.95	22	3.09	???		99.99
Total	2,739	100	713	100	≈ 560	???	≈ 100

Note. * Indicate the levels when 95% and 98% coverage were reached.

Table B.14
Tokens, types, WFs and cumulative coverage of episode G6E14 ("Cookin' with Romeo and Juliet").

Frequency band	Tokens		Types		WFs		Cumulative coverage (in %)
	Raw	%	Raw	%	Raw	%	
1k	2,215	92.37	467	75.69	373	75.51	92.37
2k	61	2.54	52	8.43	52	10.53	94.91
3k*	24	1	18	2.92	17	3.44	95.91
4k	17	0.71	15	2.43	14	2.83	96.62
5k	21	0.88	15	2.43	13	2.63	97.50
6k	10	0.42	7	1.13	7	1.42	97.92
7k*	11	0.46	4	0.65	3	0.61	98.38
8k	2	0.08	1	0.16	1	0.20	98.46
9k	2	0.08	2	0.32	2	0.40	98.54
10k	-	-	-	-	-	-	98.54
11-25k	12	0.49	12	1.93	12	2.41	99.03
Off-list	23	0.96	23	3.73	???		99.99
Total	2,398	100	617	100	≈ 494	???	≈ 100

Note. * Indicate the levels when 95% and 98% coverage were reached.

Table B.15
Tokens, types, WFs and cumulative coverage of episode G6E15 ("Rumors").

Frequency band	Tokens		Types		WFs		Cumulative coverage (in %)
	Raw	%	Raw	%	Raw	%	
1k	2,691	91.19	505	73.62	388	72.52	91.19
2k	98	3.32	70	10.20	63	11.78	94.51
3k*	31	1.05	19	2.77	18	3.36	95.56
4k	47	1.59	28	4.08	24	4.49	97.15
5k	21	0.71	18	2.62	17	3.18	97.86
6k*	16	0.54	10	1.46	10	1.87	98.40
7k	5	0.17	4	0.58	4	0.75	98.57
8k	3	0.10	2	0.29	2	0.37	98.67
9k	1	0.03	1	0.15	1	0.19	98.70
10k	3	0.10	1	0.15	1	0.19	98.80
11-25k	11	0.37	7	1.03	7	1.32	99.17
Off-list	24	0.81	20	2.92	???		99.98
Total	2,951	100	686	100	≈ 535	???	≈ 100

Note. * Indicate the levels when 95% and 98% coverage were reached.

Table B.16
Tokens, types, WFs and cumulative coverage of episode G6E16 ("Big Hair & Baseball").

Frequency band	Tokens		Types		WFs		Cumulative coverage (in %)
	Raw	%	Raw	%	Raw	%	
1k	2,534	90.02	531	69.69	402	66.45	90.02
2k	115	4.09	95	12.47	90	14.88	94.11
3k*	32	1.14	31	4.07	28	4.63	95.25
4k	29	1.03	16	2.10	15	2.48	96.28
5k	30	1.07	25	3.28	23	3.80	97.35
6k*	20	0.71	13	1.71	12	1.98	98.06
7k	8	0.28	8	1.05	8	1.32	98.34
8k	6	0.21	6	0.79	6	0.99	98.55
9k	4	0.14	4	0.52	4	0.66	98.69
10k	4	0.14	3	0.39	2	0.33	98.83
11-25k	19	0.67	16	2.09	15	2.49	99.50
Off-list	14	0.50	13	1.71	???		100
Total	2,815	100	762	100	≈ 605	???	≈ 100

Note. * Indicate the levels when 95% and 98% coverage were reached.

Table B.17
Tokens, types, WFs and cumulative coverage of episode G6E17 ("Crazy 10-Minute Sale").

Frequency band	Tokens		Types		WFs		Cumulative coverage (in %)
	Raw	%	Raw	%	Raw	%	
1k	2,861	92.08	507	77.40	387	75.88	92.08
2k*	100	3.22	57	8.70	55	10.78	95.30
3k	20	0.64	13	1.98	12	2.35	95.94
4k	14	0.45	14	2.14	14	2.75	96.39
5k	33	1.06	18	2.75	13	2.55	97.45
6k*	18	0.58	14	2.14	14	2.75	98.03
7k	4	0.13	4	0.61	4	0.78	98.16
8k	6	0.19	3	0.46	3	0.59	98.35
9k	17	0.55	2	0.31	2	0.39	98.90
10k	2	0.06	2	0.31	2	0.39	98.96
11-25k	6	0.18	4	0.60	4	0.80	99.14
Off-list	26	0.84	16	2.44	???		99.98
Total	3,107	100	655	100	≈ 510	???	≈ 100

Note. * Indicate the levels when 95% and 98% coverage were reached.

Table B.18
Tokens, types, WFs and cumulative coverage of episode G6E18 ("I Almost Drowned in a Chocolate Fountain").

Frequency band	Tokens		Types		WFs		Cumulative coverage (in %)
	Raw	%	Raw	%	Raw	%	
1k	2,521	92.01	505	77.45	384	75.74	92.01
2k*	120	4.38	66	10.12	61	12.13	96.39
3k	15	0.55	12	1.84	12	2.37	96.94
4k	18	0.66	17	2.61	16	3.16	97.60
5k*	21	0.77	14	2.15	14	2.76	98.37
6k	5	0.18	5	0.77	5	0.99	98.55
7k	7	0.26	5	0.77	5	0.99	98.81
8k	2	0.07	2	0.31	2	0.39	98.88
9k	2	0.07	2	0.31	2	0.39	98.95
10k	-	-	-	-	-	-	98.95
11-25k	6	0.23	6	0.91	6	1.19	99.18
Off-list	23	0.84	17	2.61	???		100
Total	2.740	100	652	100	≈ 507	???	≈ 100

Note. * Indicate the levels when 95% and 98% coverage were reached.

Table B.19
Tokens, types, WFs and cumulative coverage of episode G6E19 ("New Employee").

Frequency band	Tokens		Types		WFs		Cumulative coverage (in %)
	Raw	%	Raw	%	Raw	%	
1k	2,834	91.89	526	76.56	411	74.59	91.89
2k*	114	3.70	76	11.06	68	12.34	95.59
3k	28	0.91	21	3.06	19	3.45	96.50
4k	30	0.97	19	2.77	17	3.09	97.47
5k	13	0.42	9	1.31	9	1.63	97.89
6k*	18	0.58	10	1.46	10	1.81	98.47
7k	14	0.45	4	0.58	4	0.73	98.92
8k	3	0.10	3	0.44	3	0.54	99.02
9k	1	0.03	1	0.15	1	0.18	99.05
10k	2	0.06	2	0.29	2	0.36	99.11
11-25k	12	0.38	7	1.04	7	1.26	99.49
Off-list	15	0.49	9	1.31	???		99.98
Total	3,084	100	687	100	≈ 551	???	≈ 100

Note. * Indicate the levels when 95% and 98% coverage were reached.

Table B.20
Tokens, types, WFs and cumulative coverage of episode G6E20 ("Disenchanted Evening").

Frequency band	Tokens		Types		WFs		Cumulative coverage (in %)
	Raw	%	Raw	%	Raw	%	
1k	2,633	90.57	503	76.10	391	75.19	90.57
2k*	140	4.82	65	9.83	57	10.96	95.39
3k	22	0.76	16	2.42	15	2.88	96.15
4k	12	0.41	11	1.66	11	2.12	96.56
5k	16	0.55	12	1.82	12	2.31	97.11
6k	19	0.65	12	1.82	12	2.31	97.76
7k	6	0.21	6	0.91	6	1.15	97.97
8k*	7	0.24	6	0.91	5	0.96	98.21
9k	1	0.03	1	0.15	1	0.19	98.24
10k	7	0.24	4	0.61	4	0.77	98.48
11-25k	16	0.54	6	0.90	6	1.14	99.02
Off-list	28	0.96	18	2.72	???		99.98
Total	2,907	100	661	100	≈ 520	???	≈ 100

Note. * Indicate the levels when 95% and 98% coverage were reached.

Table B.21
Tokens, types, WFs and cumulative coverage of episode G6E21 ("You Can't Always Get What You Carpet").

Frequency band	Tokens		Types		WFs		Cumulative coverage (in %)
	Raw	%	Raw	%	Raw	%	
1k	2,770	92.30	468	78	357	75.80	92.30
2k*	132	4.40	71	11.83	66	14.01	96.70
3k	16	0.53	11	1.83	11	2.34	97.23
4k*	31	1.03	15	2.50	15	3.18	98.26
5k	8	0.27	7	1.17	5	1.06	98.53
6k	7	0.23	5	0.83	5	1.06	98.76
7k	4	0.13	2	0.33	2	0.42	98.89
8k	2	0.07	1	0.17	1	0.21	98.96
9k	1	0.03	1	0.17	1	0.21	98.99
10k	2	0.07	2	0.33	1	0.21	99.06
11-25k	10	0.31	7	1.19	7	1.47	99.37
Off-list	18	0.60	9	1.50	???		99.97
Total	3,001	100	600	100	≈ 471	???	≈ 100

Note. * Indicate the levels when 95% and 98% coverage were reached.

Table B.22
Tokens, types, WFs and cumulative coverage of episode G6E22 ("Curb Your Dragon").

Frequency band	Tokens		Types		WFs		Cumulative coverage (in %)
	Raw	%	Raw	%	Raw	%	
1k	2,699	91.65	489	73.20	374	72.20	91.65
2k*	115	3.90	89	13.32	81	15.64	95.55
3k	17	0.58	16	2.40	15	2.90	96.13
4k	21	0.71	13	1.95	12	2.32	96.84
5k	25	0.82	14	2.10	10	1.93	97.69
6k*	22	0.75	11	1.65	11	2.12	98.44
7k	2	0.07	2	0.30	2	0.39	98.51
8k	2	0.07	2	0.30	2	0.39	98.58
9k	3	0.10	3	0.45	3	0.58	98.68
10k	2	0.07	2	0.30	2	0.39	98.75
11-25k	12	0.40	7	1.05	6	1.15	99.15
Off-list	25	0.85	19	2.84	???		100
Total	2,945	100	668	100	≈ 518	???	≈ 100

Note. * Indicate the levels when 95% and 98% coverage were reached.

Table B.23
Tokens, types, WFs and cumulative coverage of episode G6E23 ("Pop Me and We Both Go Down").

Frequency band	Tokens		Types		WFs		Cumulative coverage (in %)
	Raw	%	Raw	%	Raw	%	
1k	2,754	89.91	532	74.93	397	72.71	89.91
2k	124	4.05	76	10.70	72	13.19	93.96
3k	12	0.39	11	1.55	11	2.01	94.35
4k*	40	1.31	17	2.39	14	2.56	95.66
5k	18	0.59	11	1.55	10	1.83	96.25
6k	17	0.56	12	1.69	12	2.20	96.81
7k	7	0.23	6	0.85	6	1.10	97.04
8k	6	0.20	5	0.70	5	0.92	97.24
9k	14	0.46	3	0.42	3	0.55	97.70
10k	6	0.20	5	0.70	5	0.92	97.90
11-25k*	22	0.71	11	1.54	11	2.01	98.61
Off-list	43	1.40	20	2.82	???		100
Total	3,063	100	710	100	≈ 546	???	≈ 100

Note. * Indicate the levels when 95% and 98% coverage were reached.

Table B.24
Tokens, types, WFs and cumulative coverage of episode G6E24 ("Potion Commotion").

Frequency band	Tokens		Types		WFs		Cumulative coverage (in %)
	Raw	%	Raw	%	Raw	%	
1k	2,789	92.11	481	75.27	368	73.02	92.11
2k*	92	3.04	67	10.49	61	12.10	95.15
3k	37	1.22	23	3.60	20	3.97	96.37
4k	13	0.43	11	1.72	11	2.18	96.80
5k	10	0.33	10	1.56	10	1.98	97.13
6k	12	0.40	10	1.56	9	1.79	97.53
7k	10	0.33	5	0.78	5	0.99	97.86
8k	-	-	-	-	-	-	97.86
9k	3	0.10	2	0.31	2	0.40	97.96
10k*	21	0.69	5	0.78	4	0.79	98.65
11-25k	23	0.77	15	2.36	14	2.79	99.42
Off-list	18	0.59	9	1.41	???		100
Total	3,028	100	639	100	≈ 504	???	≈ 100

Note. * Indicate the levels when 95% and 98% coverage were reached.

APPENDIX C - VOCABULARY DEMANDS OF TV SERIES IN GRADE 10 AND UNIVERSITY

Tables C.1-C.22 show the vocabulary demands of the 22 episodes selected in Grade 10 and the 8 episodes shown at university. They indicate the raw number and the percentage of tokens, types and WFs found in each frequency band as well as the cumulative coverage in percentage.

Table C.1
Tokens, types, WFs and cumulative coverage of episode G10E01 ("Lucy Visits Grauman's").

Frequency band	Tokens		Types		WFs		Cumulative coverage (in %)
	Raw	%	Raw	%	Raw	%	
1k	3,147	93.97	520	81.12	417	80.04	93.97
2k*	80	2.39	42	6.55	41	7.87	96.36
3k	11	0.33	10	1.56	10	1.92	96.69
4k	22	0.66	15	2.34	15	2.88	97.35
5k*	26	0.78	10	1.56	10	1.92	98.13
6k	11	0.33	7	1.09	6	1.15	98.46
7k	9	0.27	7	1.09	6	1.15	98.73
8k	1	0.03	1	0.16	1	0.19	98.76
9k	6	0.18	5	0.78	5	0.96	98.94
10k	4	0.12	2	0.31	2	0.38	99.06
11-25k	10	0.30	8	1.25	8	1.53	99.36
Off-list	22	0.66	13	2.03	???		100
Total	3,349	100	641	100	≈ 521	???	≈ 100

Note. * Indicate the levels when 95% and 98% coverage were reached.

Table C.2
Tokens, types, WFs and cumulative coverage of episode G10E02 ("Lucy and John Wayne").

Frequency band	Tokens		Types		WFs		Cumulative coverage (in %)
	Raw	%	Raw	%	Raw	%	
1k	2,958	92.90	479	77.38	379	77.19	92.90
2k*	95	2.98	65	10.50	59	12.02	95.88
3k	16	0.50	8	1.29	8	1.63	96.38
4k	18	0.57	8	1.29	7	1.43	96.95
5k	12	0.38	7	1.13	7	1.43	97.33
6k	12	0.38	11	1.78	10	2.04	97.71
7k	8	0.25	4	0.65	4	0.81	97.96
8k*	4	0.13	4	0.65	3	0.61	98.09
9k	1	0.03	1	0.16	1	0.20	98.12
10k	4	0.13	3	0.48	3	0.61	98.25
11-25k	18	0.55	10	1.60	10	2.01	98.80
Off-list	38	1.19	18	2.91	???		99.99
Total	3,184	100	619	100	≈ 491	???	≈ 100

Note. * Indicate the levels when 95% and 98% coverage were reached.

Table C.3
Tokens, types, WFs and cumulative coverage of episode G10E03 ("Lucy and the Dummy").

Frequency band	Tokens		Types		WFs		Cumulative coverage (in %)
	Raw	%	Raw	%	Raw	%	
1k	2,889	94.94	475	83.92	384	82.58	94.94
2k*	78	2.56	48	8.48	47	10.11	97.50
3k*	33	1.08	13	2.30	12	2.58	98.58
4k	8	0.26	7	1.24	7	1.51	98.84
5k	-	-	-	-	-	-	98.84
6k	7	0.23	5	0.88	5	1.08	99.07
7k	4	0.13	3	0.53	3	0.65	99.20
8k	5	0.16	2	0.35	2	0.43	99.36
9k	1	0.03	1	0.18	1	0.22	99.39
10k	1	0.03	1	0.18	1	0.22	99.42
11-25k	5	0.16	3	0.54	3	0.66	99.58
Off-list	12	0.39	7	1.24	???		99.97
Total	3,042	100	566	100	≈ 465	???	≈ 100

Note. * Indicate the levels when 95% and 98% coverage were reached.

Table C.4
Tokens, types, WFs and cumulative coverage of episode G10E04 ("Ricky Sells the Car").

Frequency band	Tokens		Types		WFs		Cumulative coverage (in %)
	Raw	%	Raw	%	Raw	%	
1k	3,200	94.03	459	81.82	354	80.82	94.03
2k*	116	3.41	53	9.45	45	10.27	97.44
3k	12	0.35	7	1.25	6	1.37	97.79
4k*	9	0.26	8	1.43	8	1.83	98.05
5k	23	0.73	13	2.32	10	2.28	98.78
6k	7	0.21	4	0.71	4	0.91	98.99
7k	5	0.15	3	0.53	2	0.46	99.14
8k	1	0.03	1	0.18	1	0.23	99.17
9k	1	0.03	1	0.18	1	0.23	99.20
10k	1	0.03	1	0.18	1	0.23	99.23
11-25k	12	0.36	6	1.08	6	1.38	99.59
Off-list	14	0.41	4	0.71	???		100
Total	3,403	100	561	100	≈ 438	???	≈ 100

Note. * Indicate the levels when 95% and 98% coverage were reached.

Table C.5
Tokens, types, WFs and cumulative coverage of episode G10E05 ("The Great Train Robbery").

Frequency band	Tokens		Types		WFs		Cumulative coverage (in %)
	Raw	%	Raw	%	Raw	%	
1k	2,990	93.09	460	78.36	366	77.05	93.09
2k*	118	3.67	64	10.90	60	12.63	96.76
3k	8	0.25	6	1.02	6	1.26	97.01
4k	17	0.53	9	1.53	9	1.89	97.54
5k*	24	0.75	7	1.19	7	1.47	98.29
6k	15	0.47	11	1.87	11	2.32	98.76
7k	13	0.40	5	0.85	4	0.84	99.16
8k	1	0.03	1	0.17	1	0.21	99.19
9k	4	0.12	3	0.51	3	0.63	99.31
10k	2	0.06	2	0.34	2	0.42	99.37
11-25k	6	0.18	6	1.02	6	1.26	99.55
Off-list	14	0.44	12	2.04	???		99.99
Total	3,212	100	587	100	≈ 475	???	≈ 100

Note. * Indicate the levels when 95% and 98% coverage were reached.

Table C.6
Tokens, types, WFs and cumulative coverage of episode G10E06 ("Homecoming").

Frequency band	Tokens		Types		WFs		Cumulative coverage (in %)
	Raw	%	Raw	%	Raw	%	
1k	2,978	94.51	523	81.21	412	80.16	94.51
2k*	79	2.51	49	7.61	44	8.56	97.02
3k	15	0.48	14	2.17	12	2.33	97.50
4k*	19	0.60	13	2.02	12	2.33	98.10
5k	14	0.44	10	1.55	10	1.95	98.54
6k	9	0.29	9	1.40	8	1.56	98.83
7k	3	0.10	2	0.31	2	0.39	98.93
8k	5	0.16	4	0.62	4	0.78	99.09
9k	4	0.13	3	0.47	3	0.58	99.22
10k	1	0.03	1	0.16	1	0.19	99.25
11-25k	8	0.24	7	1.11	6	1.14	99.49
Off-list	16	0.51	8	1.24	???		100
Total	3,151	100	644	100	≈ 514	???	≈ 100

Note. * Indicate the levels when 95% and 98% coverage were reached.

Table C.7
Tokens, types, WFs and cumulative coverage of episode G10E07 ("Face to Face").

Frequency band	Tokens		Types		WFs		Cumulative coverage (in %)
	Raw	%	Raw	%	Raw	%	
1k	3,681	94.72	571	80.65	449	79.89	94.72
2k*	112	2.88	64	9.04	59	10.50	97.60
3k*	21	0.54	15	2.12	15	2.67	98.14
4k	15	0.39	12	1.69	10	1.78	98.53
5k	13	0.33	13	1.84	13	2.31	98.86
6k	10	0.26	6	0.85	6	1.07	99.12
7k	1	0.03	1	0.14	1	0.18	99.15
8k	2	0.05	2	0.28	2	0.36	99.20
9k	1	0.03	1	0.14	1	0.18	99.23
10k	5	0.13	2	0.28	1	0.18	99.36
11-25k	10	0.26	6	0.84	5	0.90	99.62
Off-list	15	0.39	15	2.12	???		100
Total	3,886	100	708	100	≈ 562	???	≈ 100

Note. * Indicate the levels when 95% and 98% coverage were reached.

Table C.8
Tokens, types, WFs and cumulative coverage of episode G10E08 ("Nursery School").

Frequency band	Tokens		Types		WFs		Cumulative coverage (in %)
	Raw	%	Raw	%	Raw	%	
1k	3,176	91.87	504	77.18	390	75.14	91.87
2k*	124	3.59	70	10.72	64	12.33	95.46
3k	22	0.64	17	2.60	17	3.28	96.10
4k	44	1.27	8	1.23	7	1.35	97.37
5k	10	0.29	6	0.92	5	0.96	97.66
6k	9	0.26	7	1.07	7	1.35	97.92
7k*	15	0.43	6	0.92	6	1.16	98.35
8k	7	0.20	4	0.61	4	0.77	98.55
9k	6	0.17	4	0.61	4	0.77	98.72
10k	9	0.26	5	0.77	4	0.77	98.98
11-25k	23	0.68	12	1.83	12	2.31	99.66
Off-list	12	0.35	9	1.38	???		100
Total	3,457	100	653	100	≈ 519	???	≈ 100

Note. * Indicate the levels when 95% and 98% coverage were reached.

Table C.9
Tokens, types, WFs and cumulative coverage of episode G10E09 ("The Trip (Part 1)").

Frequency band	Tokens		Types		WFs		Cumulative coverage (in %)
	Raw	%	Raw	%	Raw	%	
1k	2,908	91.42	574	73.87	456	72.15	91.42
2k*	125	3.93	96	12.36	87	13.77	95.35
3k	31	0.97	25	3.22	24	3.80	96.32
4k	44	1.38	25	3.22	21	3.32	97.70
5k*	12	0.38	11	1.42	11	1.74	98.08
6k	9	0.28	7	0.90	7	1.11	98.36
7k	7	0.22	7	0.90	7	1.11	98.58
8k	8	0.25	6	0.77	6	0.95	98.83
9k	8	0.25	4	0.51	4	0.63	99.08
10k	4	0.13	3	0.39	3	0.47	99.21
11-25k	8	0.25	6	0.78	6	0.96	99.46
Off-list	17	0.53	12	1.54	???		99.99
Total	3,181	100	777	100	≈ 632	???	≈ 100

Note. * Indicate the levels when 95% and 98% coverage were reached.

Table C.10
Tokens, types, WFs and cumulative coverage of episode G10E10 ("The Trip (Part 2)").

Frequency band	Tokens		Types		WFs		Cumulative coverage (in %)
	Raw	%	Raw	%	Raw	%	
1k	2,837	91.90	530	73.20	412	72.03	91.90
2k*	115	3.73	91	12.57	80	13.99	95.63
3k	24	0.78	20	2.76	19	3.32	96.41
4k	31	1	19	2.62	15	2.62	97.41
5k*	28	0.91	21	2.90	18	3.15	98.32
6k	15	0.49	13	1.80	10	1.75	98.81
7k	6	0.19	6	0.83	6	1.05	99.00
8k	3	0.10	3	0.41	3	0.52	99.10
9k	3	0.10	3	0.41	3	0.52	99.20
10k	3	0.10	3	0.41	3	0.52	99.30
11-25k	9	0.29	5	0.70	5	0.87	99.59
Off-list	13	0.42	11	1.52	???		100
Total	3,087	100	724	100	≈ 572	???	≈ 100

Note. * Indicate the levels when 95% and 98% coverage were reached.

Table C.11
Tokens, types, WFs and cumulative coverage of episode G10E11 ("The Bubble Boy").

Frequency band	Tokens		Types		WFs		Cumulative coverage (in %)
	Raw	%	Raw	%	Raw	%	
1k	2,903	91.95	509	78.43	399	77.78	91.95
2k	81	2.57	59	9.09	53	10.33	94.52
3k*	20	0.63	15	2.31	15	2.92	95.15
4k	66	2.09	16	2.47	16	3.12	97.24
5k	11	0.35	8	1.23	7	1.36	97.59
6k	8	0.25	8	1.23	7	1.36	97.84
7k*	6	0.19	5	0.77	5	0.97	98.03
8k	-	-	-	-	-	-	98.03
9k	1	0.03	1	0.15	1	0.19	98.06
10k	-	-	-	-	-	-	98.06
11-25k*	14	0.44	10	1.53	10	1.93	98.50
Off-list	47	1.49	17	2.62	???		99.99
Total	3,157	100	649	100	≈ 513	???	≈ 100

Note. * Indicate the levels when 95% and 98% coverage were reached.

Table C.12
Tokens, types, WFs and cumulative coverage of episode G10E12 ("The Cheever Letters").

Frequency band	Tokens		Types		WFs		Cumulative coverage (in %)
	Raw	%	Raw	%	Raw	%	
1k	2,929	94.03	518	79.33	404	77.54	94.03
2k*	75	2.41	63	9.65	61	11.71	96.44
3k	22	0.71	19	2.91	19	3.65	97.15
4k*	29	0.93	12	1.84	12	2.30	98.08
5k	25	0.80	10	1.53	7	1.73	98.88
6k	3	0.10	3	0.46	3	0.58	98.98
7k	5	0.16	5	0.77	5	0.96	99.14
8k	1	0.03	1	0.15	1	0.19	99.17
9k	1	0.03	1	0.15	1	0.19	99.20
10k	5	0.16	1	0.15	1	0.19	99.36
11-25k	6	0.18	6	0.91	5	0.95	99.54
Off-list	14	0.45	13	1.99	???		99.99
Total	3,115	100	653	100	≈ 521	???	≈ 100

Note. * Indicate the levels when 95% and 98% coverage were reached.

Table C.13
Tokens, types, WFs and cumulative coverage of episode G10E13 ("The Airport").

Frequency band	Tokens		Types		WFs		Cumulative coverage (in %)
	Raw	%	Raw	%	Raw	%	
1k	2,756	90.45	503	73.22	393	71.07	90.45
2k*	166	5.45	91	13.25	88	15.91	95.90
3k	22	0.72	19	2.77	19	3.44	96.62
4k	15	0.49	14	2.04	13	2.35	97.11
5k	18	0.59	15	2.18	15	2.71	97.70
6k*	9	0.30	9	1.31	7	1.27	98.00
7k	5	0.16	3	0.44	3	0.54	98.16
8k	3	0.10	3	0.44	3	0.54	98.26
9k	6	0.20	4	0.58	4	0.72	98.46
10k	12	0.39	4	0.58	4	0.72	98.85
11-25k	8	0.25	4	0.60	4	0.72	99.10
Off-list	27	0.89	17	2.47	???		99.99
Total	3,047	100	687	100	≈ 553	???	≈ 100

Note. * Indicate the levels when 95% and 98% coverage were reached.

Table C.14
Tokens, types, WFs and cumulative coverage of episode G10E14 ("The Movie").

Frequency band	Tokens		Types		WFs		Cumulative coverage (in %)
	Raw	%	Raw	%	Raw	%	
1k	3,122	94.09	526	81.42	411	80.43	94.09
2k*	85	2.56	55	8.51	45	8.81	96.65
3k	29	0.87	18	2.79	17	3.33	97.52
4k	5	0.15	4	0.62	4	0.78	97.67
5k*	15	0.45	12	1.86	11	2.15	98.12
6k	20	0.60	10	1.55	10	1.96	98.72
7k	14	0.42	4	0.62	4	0.78	99.14
8k	5	0.15	3	0.46	3	0.59	99.29
9k	2	0.06	1	0.15	1	0.20	99.35
10k	2	0.06	2	0.31	2	0.39	99.41
11-25k	3	0.09	3	0.45	3	0.60	99.50
Off-list	16	0.48	7	1.08	???		99.98
Total	3,318	100	646	100	≈ 511	???	≈ 100

Note. * Indicate the levels when 95% and 98% coverage were reached.

Table C.15
Tokens, types, WFs and cumulative coverage of episode G10E15 ("The Shoes").

Frequency band	Tokens		Types		WFs		Cumulative coverage (in %)
	Raw	%	Raw	%	Raw	%	
1k	2,925	92.68	501	78.04	384	76.04	92.68
2k	70	2.22	58	9.03	57	11.29	94.90
3k*	37	1.17	24	3.74	21	4.16	96.07
4k	31	0.98	17	2.65	16	3.17	97.05
5k	18	0.57	9	1.40	9	1.78	97.62
6k*	19	0.60	7	1.09	7	1.39	98.22
7k	4	0.13	3	0.47	2	0.40	98.35
8k	9	0.29	4	0.62	3	0.59	98.64
9k	16	0.51	5	0.78	2	0.40	99.15
10k	5	0.16	1	0.16	1	0.20	99.31
11-25k	12	0.38	3	0.48	3	0.60	99.69
Off-list	10	0.32	9	1.40	???		100
Total	3,156	100	642	100	≈ 505	???	≈ 100

Note. * Indicate the levels when 95% and 98% coverage were reached.

Table C.16
Tokens, types, WFs and cumulative coverage of episode G10E16 ("The Glasses").

Frequency band	Tokens		Types		WFs		Cumulative coverage (in %)
	Raw	%	Raw	%	Raw	%	
1k	2,602	92.83	473	76.54	371	75.10	92.83
2k*	88	3.14	60	9.71	53	10.73	95.97
3k	27	0.96	21	3.40	19	3.85	96.93
4k	22	0.78	14	2.27	14	2.83	97.71
5k*	16	0.57	12	1.94	12	2.43	98.28
6k	17	0.61	11	1.78	8	1.62	98.89
7k	4	0.14	4	0.65	4	0.81	99.03
8k	4	0.14	4	0.65	4	0.81	99.17
9k	-	-	-	-	-	-	99.17
10k	5	0.18	4	0.65	3	0.61	99.35
11-25k*	10	0.36	6	0.96	6	1.20	99.71
Off-list	8	0.29	8	1.29	???		100
Total	2,803	100	618	100	≈ 494	???	≈ 100

Note. * Indicate the levels when 95% and 98% coverage were reached.

Table C.17
Tokens, types, WFs and cumulative coverage of episode G10E17 (“The Sniffing Accountant”).

Frequency band	Tokens		Types		WFs		Cumulative coverage (in %)
	Raw	%	Raw	%	Raw	%	
1k	2,826	91.49	513	75.22	397	72.58	91.49
2k	91	2.95	72	10.56	68	12.43	94.44
3k*	25	0.81	20	2.93	20	3.66	95.25
4k	45	1.46	22	3.23	22	4.02	96.71
5k	26	0.84	8	1.17	7	1.28	97.55
6k*	24	0.78	10	1.47	6	1.10	98.33
7k	5	0.16	5	0.73	5	0.91	98.49
8k	8	0.26	7	1.03	7	1.28	98.75
9k	3	0.10	3	0.44	4	0.55	98.85
10k	6	0.19	3	0.44	3	0.55	99.04
11-25k	13	0.41	10	1.48	9	1.64	99.45
Off-list	17	0.55	8	1.17	???		100
Total	3,089	100	682	100	≈ 547	???	≈ 100

Note. * Indicate the levels when 95% and 98% coverage were reached.

Table C.18
Tokens, types, WFs and cumulative coverage of episode G10E18 (“The Bris”).

Frequency band	Tokens		Types		WFs		Cumulative coverage (in %)
	Raw	%	Raw	%	Raw	%	
1k	3,109	90.06	571	68.22	440	65.57	90.06
2k	146	4.23	107	12.78	98	14.61	94.29
3k*	51	1.48	42	5.02	38	5.66	95.77
4k	23	0.67	21	2.51	20	2.98	96.44
5k	20	0.58	18	2.15	17	2.53	97.02
6k	11	0.32	11	1.31	10	1.49	97.34
7k*	24	0.70	18	2.15	16	2.38	98.04
8k	14	0.41	8	0.96	7	1.04	98.45
9k	8	0.23	6	0.72	5	0.75	98.68
10k	5	0.14	5	0.60	5	0.75	98.82
11-25k	27	0.79	15	1.80	15	2.25	99.61
Off-list	14	0.41	14	1.67	???		100
Total	3,452	100	837	100	≈ 671	???	≈ 100

Note. * Indicate the levels when 95% and 98% coverage were reached.

Table C.19
Tokens, types, WFs and cumulative coverage of episode G10E19 ("The Lip Reader").

Frequency band	Tokens		Types		WFs		Cumulative coverage (in %)
	Raw	%	Raw	%	Raw	%	
1k	2,918	92.96	532	77.21	413	75.50	92.26
2k*	88	2.80	68	9.87	61	11.15	95.76
3k	34	1.08	25	3.63	23	4.20	96.84
4k	25	0.80	13	1.89	13	2.38	97.64
5k	11	0.35	9	1.31	8	1.46	97.99
6k*	14	0.45	9	1.31	9	1.65	98.44
7k	6	0.19	5	0.73	4	0.73	98.63
8k	7	0.22	5	0.73	5	0.91	98.85
9k	6	0.19	3	0.44	2	0.37	99.04
10k	2	0.06	2	0.29	2	0.37	99.10
11-25k	9	0.28	7	1.04	7	1.27	99.38
Off-list	19	0.61	10	1.45	???		99.99
Total	3,139	100	689	100	≈ 547	???	≈ 100

Note. * Indicate the levels when 95% and 98% coverage were reached.

Table C.20
Tokens, types, WFs and cumulative coverage of episode G10E20 ("The Non-Fat Yogurt").

Frequency band	Tokens		Types		WFs		Cumulative coverage (in %)
	Raw	%	Raw	%	Raw	%	
1k	2,712	91.47	510	74.45	394	73.64	91.47
2k	104	3.51	80	11.68	70	13.08	94.98
3k*	26	0.88	23	3.36	21	3.93	95.86
4k	32	1.08	26	3.80	24	4.49	96.94
5k	6	0.20	5	0.73	5	0.93	97.14
6k	9	0.30	6	0.88	5	0.93	97.44
7k*	23	0.78	3	0.44	3	0.56	98.22
8k	6	0.20	5	0.73	4	0.75	98.42
9k	3	0.10	3	0.44	3	0.56	98.52
10k	1	0.03	1	0.15	1	0.19	98.55
11-25k	6	0.20	5	0.74	5	0.94	98.75
Off-list	37	1.25	17	2.48	???		100
Total	2,965	100	685	100	≈ 535	???	≈ 100

Note. * Indicate the levels when 95% and 98% coverage were reached.

Table C.21
Tokens, types, WFs and cumulative coverage of episode G10E21 ("The Conversion").

Frequency band	Tokens		Types		WFs		Cumulative coverage (in %)
	Raw	%	Raw	%	Raw	%	
1k	2,324	90.60	472	74.68	350	70.99	90.60
2k	88	3.43	65	10.28	62	12.58	94.03
3k*	38	1.48	25	3.96	24	4.87	95.51
4k	15	0.58	11	1.74	11	2.23	96.09
5k	34	1.33	18	2.85	14	2.84	97.42
6k*	17	0.66	8	1.27	8	1.62	98.08
7k	4	0.16	4	0.63	4	0.81	98.24
8k	6	0.23	5	0.79	5	1.01	98.47
9k	1	0.04	1	0.16	1	0.20	98.51
10k*	6	0.23	5	0.79	5	1.01	98.74
11-25k	21	0.82	11	1.74	9	1.83	99.56
Off-list	11	0.43	6	0.95	???		99.99
Total	2,565	100	632	100	≈ 493	???	≈ 100

Note. * Indicate the levels when 95% and 98% coverage were reached.

Table C.22
Tokens, types, WFs and cumulative coverage of episode G10E22 ("The Stall").

Frequency band	Tokens		Types		WFs		Cumulative coverage (in %)
	Raw	%	Raw	%	Raw	%	
1k	2,721	91.28	455	73.03	355	70.72	91.28
2k	105	3.52	60	9.63	56	11.16	94.80
3k*	21	0.70	20	3.21	20	3.98	95.50
4k	34	1.14	26	4.17	23	4.58	96.64
5k	14	0.47	8	1.28	7	1.39	97.11
6k	22	0.74	15	2.41	15	2.99	97.85
7k*	7	0.23	6	0.96	6	1.20	98.08
8k	12	0.40	3	0.48	3	0.60	98.48
9k	11	0.37	6	0.96	6	1.20	98.85
10k	2	0.07	2	0.32	2	0.40	98.92
11-25k	12	0.39	9	1.44	9	1.80	99.31
Off-list	20	0.67	12	1.93	???		99.98
Total	2,981	100	623	100	≈ 502	???	≈ 100

Note. * Indicate the levels when 95% and 98% coverage were reached.

APPENDIX D - LIST OF TWs IN GRADE 6

Tables D.1-D.3 show the features of the 120 TWs which were selected in Grade 6.

Table D.1
Features of the TWs selected in Grade 6 in Term 1 (*The Suite Life of Zack and Cody*).

TW	Frequency				PoS	Concreteness (Mean)	Cognateness (Y / N)
	Episode (*N* times)	Term (*N* times)	COCA (Freq. band)	$SUBTLEX_{US}$ (%)			
Admirer	9	9	2k	1.48	N	2.93	N
Ankle	3	3	4k	3.41	N	4.81	N
To apologize	2	2	3k	20.09	V	2.63	N
Awesome	2	4	6k	10.41	Adj.	1.83	N
Badge	2	2	5k	5.52	N	4.93	N
Ballroom	3	6	Off-list	1.50	N	4.70	N
Beauty mark*	2	2	1k	17.93	N	4.21	N
Burden	4	4	3k	4.66	N	2.63	N
Concierge	3	3	12k	0.60	N	3.89	Y
Contest	7	9	3k	5.96	N	3.52	N
Counter	2	4	2k	6.44	N	4.17	N
Cute	5	11	6k	29.32	Adj.	2.76	N
Date	5	17	1k	35.61	N	3.90	N
To ditch	5	5	5k	3.67	V	4.50	N
Dweeb	4	4	20k	0.25	N	2.77	N
Goat	2	2	4k	4.03	N	5	N
Grounded	5	6	1k	3.24	Adj.	2.07	N
To hurt	6	11	1k	61.08	V	3.61	N
Lame	5	8	7k	5.23	Adj.	2.52	N
Lifeguard	4	4	Off-list	0.66	N	4.47	N
Lipstick	2	2	Off-list	3.95	N	4.90	N
Lobby	6	10	3k	5.40	N	4.70	N
To owe	2	3	2k	28.05	V	2.07	N
Pageant	11	11	7k	0.85	N	4.00	N
Prize	2	2	2k	8.76	N	4.45	N
Prom	23	23	9k	3.83	N	3.72	N
Review	4	5	3k	6.33	N	2.81	N
Slumber party*	2	2	9k	1.25	N	3.89	N
Step	4	6	1k	39.85	N	4.54	N
Sweet	4	16	1k	42.66	Adj.	4	N
Terrific	2	2	5k	15.19	Adj.	2.07	N
Thirsty	3	3	1k	6.04	Adj.	3.86	N
Tux	3	3	Off-list	1.48	N	4.96	N
Vase	4	5	5k	1.44	N	5	N
Vent	6	6	5k	1.90	N	4.56	Y
Wedding	9	9	1k	19.72	N	3.92	N
Weenies	4	4	17k	0.24	N	4.22	N
To work out*	3	3	1k	91.44	V	3.48	N
Wrestler	3	3	5k	0.69	N	4.42	N
To yell	4	5	2k	8.71	V	3.86	N

Note. * In the case of compound nouns and phrasal verbs, only the $SUBTLEX_{US}$ percentage and the concreteness mean of the lexical verb (for phrasal verbs) or the less frequent word (for compound nouns) is reported.

Table D.2
Features of the TWs selected in Grade 6 in Term 2 (*The Suite Life of Zack and Cody*).

TW	Frequency				PoS	Concreteness (Mean)	Cognateness (Y / N)
	Episode (*N* times)	Term (*N* times)	COCA (Freq. band)	$SUBTLEX_{US}$ (%)			
Beneath	4	5	1k	5.54	Prep.	3.11	N
Birth certificate*	3	3	4k	3.54	N	5	Y
To blot	2	2	6k	0.44	V	3.44	N
Bossy	5	5	2k	0.81	Adj.	2.21	N
To care	4	19	1k	82.80	V	2.33	N
Chickenpox	2	2	Off-list	0.13	N	4.56	N
Cookie	9	9	7k	5.34	N	5	N
Crook	7	7	5k	2.34	N	3.79	N
Diaper	3	4	10k	1.91	N	4.82	N
Dungeon	3	3	10k	1.19	N	4.32	N
To eavesdrop	2	2	9k	0.46	V	2.85	N
To fast forward*	5	5	1k	27.91	V	2.66	N
To fix	6	10	1k	30.71	V	2.93	N
To flick	3	3	4k	1.63	V	3.71	N
To flunk	2	2	13k	0.89	V	2.40	N
Footprint	3	3	2k	0.55	N	4.37	N
Fox	4	5	2k	4.88	N	4.97	N
Frizzy	3	3	12k	0.13	Adj.	3.83	N
Gearshift	2	2	Off-list	0.13	N	4.64	N
To gossip	9	9	4k	3.36	V	1.54	N
Jewel	13	14	5k	2.15	N	4.96	N
To lock	4	5	1k	22.25	V	4.65	N
Loot	3	3	5k	1.62	N	4.50	Y
Lounge	5	7	10k	3.25	N	3.96	N
To mix up	3	3	2k	8.07	V	3.82	N
Nervous wreck*	2	2	4k	6.18	N	1.85	Y
Outfit	3	6	4k	9.99	N	4.12	N
Peanut	6	10	6k	4.32	N	4.89	N
Pie	4	6	2k	9.42	N	4.92	N
Pillow	2	4	4k	5.40	N	5	N
To pledge	3	3	4k	2.75	V	2.61	N
Recess	4	4	6k	1.51	N	3.07	N
To rehearse	4	4	4k	1.99	V	3.32	N
Smart	5	9	2k	33.30	Adj.	1.75	N
Spy gear	4	4	2k	6.14	N	4.14	Y
To stick	6	9	1k	35.71	V	4.59	N
To sweat	4	4	3k	10.29	V	4.71	Y
Trap	4	5	2k	9.61	N	4.30	Y
Twin	7	11	2k	5.65	N	4.17	N
Whistle	5	5	2k	5.64	N	4.42	N

Note. * In the case of compound nouns and phrasal verbs, only the $SUBTLEX_{US}$ percentage and the concreteness mean of the lexical verb (for phrasal verbs) or the less frequent word (for compound nouns) is reported.

Table D.3
Features of the TWs selected in Grade 6 in Term 3 (*Wizards of Waverly Place*).

TW	Frequency				PoS	Concreteness (Mean)	Cognateness (Y / N)
	Episode (*N* times)	**Term** (*N* times)	**COCA** (Freq. band)	**SUBTLEX$_{US}$** (%)			
Ant	5	5	5k	1.32	N	4.86	N
Apron	3	3	6k	1.28	N	4.87	N
Backpack	2	2	Off-list	1.56	N	4.96	N
To bark	2	4	2k	2.47	V	4.52	N
Basement	5	6	4k	8.09	N	4.89	N
Brick	11	11	2k	3.04	N	4.83	Y
Cage	3	3	2k	6.72	N	5	N
Carpet	12	12	2k	4.58	N	4.96	N
Charm	6	8	2k	7.12	N	2.76	N
To cheat	6	11	2k	7.34	V	2.23	N
Commitment	3	3	2k	5.22	N	1.74	Y
To curl	5	5	2k	1.32	V	4.38	N
Dandruff	2	2	13k	0.38	N	4.79	N
Drill	5	5	3k	5.56	N	4.40	N
Elf	28	28	5k	0.91	N	4.30	Y
Essay	3	4	3k	1.53	N	3.97	N
To fail	14	15	2k	11.19	V	1.44	N
To fire	14	14	1k	47.07	V	2.48	N
Fur	4	6	2k	3.40	N	4.69	N
Freewheeling	3	3	Off-list	0.06	Adj.	-	N
Goth	5	5	Off-list	0.30	Adj.	2.83	Y
To hire	5	5	2k	11.40	V	3.11	N
Jail	3	4	3k	20.33	N	4.83	N
Lair	3	3	10k	0.88	N	3.76	N
Owner	9	12	1k	10.13	N	3.78	N
Partner	4	4	2k	21.13	N	3.53	Y
To pop	5	9	1k	18.45	V	4.22	N
Potion	17	18	10k	1.34	N	4.38	Y
Purse	16	18	5k	8.01	N	4.90	N
Rightful	4	4	1k	1.28	Adj.	1.60	N
To schmooze	5	5	15k	0.18	V	2.29	N
Sneakers	4	4	5k	1.42	N	4.69	N
Spell	8	47	2k	10.54	N	3.32	N
Summit	12	12	3k	0.94	N	4.21	N
Unsightly	5	5	1k	0.21	Adv.	2.11	N
Waitress	5	5	4k	4.74	N	4.56	N
Wallpaper	9	9	Off-list	1.45	N	4.62	N
Wand	16	16	9k	0.88	N	4.73	N
Wizard	13	29	5k	2.59	N	4.43	N
Zit	12	12	18k	0.48	N	4.85	N

APPENDIX E - LIST OF TWs IN GRADE 10 AND UNIVERSITY

Tables E.1-E.3 show the features of the 110 TWs which were selected in Grade 10. Table E.1 also describes the TWs that were tested at university in Term 1.

Table E.1
Features of the TWs selected in Grade 10 and university in Term 1 (*I Love Lucy*).

TW	Frequency				PoS	Concreteness (Mean)	Cognateness (Y / N)
	Episode (*N* times)	Term (*N* times)	COCA (Freq. band)	SUBTLEX$_{US}$ (%)			
Beady eyes*	2	2	4k	0.33	N	3.46	N
Bucket	5	5	2k	4.66	N	4.96	N
Cheapskate	2	2	17k	0.30	Adj.	3.26	N
Choppy	3	3	11k	0.15	Adj.	2.92	N
Conductor	4	4	3k	1.28	N	4.26	N
Crowbar	3	5	12k	0.67	N	4.87	N
Crummy	4	4	14k	1.67	Adj.	2.57	N
Cue tip*	2	2	5k	3.49	N	3.31	N
Cuffs	5	5	6k	2.36	N	4.61	N
Curler	4	4	2k	0.06	N	4.13	N
Downtown	4	4	Off-list	9.75	N	4.39	N
Dummy	3	3	6k	3.98	N	3.96	N
To fool	3	5	2k	28.25	V	3.19	N
Forecourt	6	6	13k	0.02	N	-	N
To forge	4	4	4k	1.23	V	4.04	N
Furnace	2	3	7k	1.05	N	4.69	N
Galoshes	4	4	16k	0.27	N	-	N
Gear	2	2	2k	6.66	N	4.28	N
Grapefruit	2	3	9k	0.99	N	4.96	N
To hobnob	2	2	16k	0.11	V	-	N
Hunk	2	3	9k	2.43	N	3.79	N
Mitten	3	3	10k	0.45	N	4.89	N
Newsstand	2	2	Off-list	0.50	N	4.71	N
To peek through*	5	5	7k	2.69	V	3.62	N
Penthouse	5	5	10k	1.80	N	4.36	N
Preview	4	7	6k	0.75	N	3.20	Y
Razor	3	3	6k	3	N	4.90	N
To roll	6	7	1k	22.82	V	4.16	Y
Rubdown	4	4	Off-list	0.30	N	3.62	N
Rugged	3	3	6k	1.05	Adj.	2.75	N
Seasick	2	2	Off-list	0.62	Adj.	3.18	N
Sneaky	2	3	5k	2.19	Adj.	1.97	N
To sneeze	2	2	8k	1.23	V	4.03	N
To snore	3	3	7k	0.86	V	4.39	N
Tenant	4	4	4k	1.20	N	4.55	N
Tonsils	6	6	12k	0.76	N	4.72	N
To totter	4	4	10k	0.02	V	-	N
To tuck in	2	2	4k	3.12	V	3.86	N
Upper berth*	4	4	7k	0.37	N	3.37	N
To vacuum	2	2	5k	2.48	V	4.22	N

Note. * In the case of compound nouns and phrasal verbs, only the SUBTLEX$_{US}$ percentage and the concreteness mean of the lexical verb (for phrasal verbs) or the less frequent word (for compound nouns) is reported.

Table E.2
Features of the TWs selected in Grade 10 in Term 2 (*Seinfeld*).

TW	Frequency				PoS	Concreteness (Mean)	Cognateness (Y / N)
	Episode (*N* times)	**Term** (*N* times)	**COCA** (Freq. band)	**SUBTLEX$_{US}$** (%)			
Anklet	3	3	16k	0.08	N	4.75	N
Awning	4	4	10k	0.21	N	4.61	N
Blanket	6	7	2k	5.72	N	5	N
Blurb	5	5	13k	0.06	N	3.04	N
To break	2	3	1k	59.88	V	3.71	N
Butler	6	6	8k	2.18	N	4.63	N
Cabin	11	27	4k	5.47	N	4.92	Y
To cherish	2	2	6k	1.93	V	1.73	N
To choke	2	3	4k	3.24	V	3.97	N
Cigar	16	20	5k	4.61	N	4.93	Y
Cleavage	13	13	6k	0.72	N	4.64	N
Coach	4	4	2k	7.14	N	4.12	N
Conveyor belt*	2	2	3k	0.41	N	4.90	N
Dill	3	3	10k	0.39	N	4.07	N
Filthy	2	3	5k	7.08	Adj.	3.53	N
Gauge	6	6	4k	1.10	N	4	N
To leer	4	4	9k	0.08	V	3.75	N
Lieutenant	5	6	4k	10.56	N	4.41	Y
Narc	3	3	15k	0.51	N	-	Y
Nostrils	3	3	6k	0.95	N	4.89	N
Nut	3	3	2k	6.57	N	4.52	N
Pickup	4	4	5k	3.48	N	4.24	N
To poke	6	6	4k	2.73	V	4.22	N
Prude	2	2	12k	0.69	N	2.39	N
Remote	4	4	3k	5.96	N	4.93	Y
To rip	3	3	2k	9.12	V	3.79	N
Script	2	13	4k	4.82	N	4.72	N
Shackles	2	2	9k	0.49	N	4.34	N
To snub	12	12	9k	0.19	V	2.41	N
Spot	8	8	1k	24.19	N	4.21	N
Stub	11	11	7k	0.41	N	4.40	N
Toll	2	3	5k	1.80	N	3.54	N
Treatment	4	6	1k	7.68	N	2.93	N
Weed	2	2	2k	3.41	N	4.90	N
To wipe	2	2	2k	8.64	V	4	N

Note. * In the case of compound nouns, only the SUBTLEX$_{US}$ percentage and the concreteness mean of the less frequent word is reported.

Table E.3
Features of the TWs selected in Grade 10 in Term 3 (*Seinfeld*).

TW	Frequency				PoS	Concreteness (Mean)	Cognateness (Y / N)
	Episode (*N* times)	Term (*N* times)	COCA (Freq. band)	SUBTLEX$_{US}$ (%)			
Accountant	5	5	4k	2.28	N	4.41	Y
Brassiere	4	4	6k	0.29	N	4.64	N
To curse	5	5	4k	6.16	V	1.96	N
Deaf	12	12	4k	5.75	Adj.	3.27	N
To flinch	7	7	7k	0.48	V	3.59	N
Flinty	4	4	9k	0.05	Adj.	-	N
Frames	7	7	2k	1	N	4.30	N
Fungus	8	8	5k	0.92	N	4.59	Y
Godfather	7	7	6k	1.45	N	4.14	N
Lineswoman	8	8	Off-list	-	N	-	N
Lobster	8	8	6k	2.40	N	4.86	N
Loop	3	3	4k	3.04	N	4.26	N
Lure	3	3	5k	2.24	N	3.59	N
Mayor	9	9	3k	6.06	N	4.37	N
Medicine cabinet*	6	6	3k	3.72	N	4.89	N
To melt	2	2	2k	3.18	V	3.96	N
Mohair	4	4	13k	0.11	N	-	N
Nametag	9	9	Off-list	-	N	4.73	N
Napkin	3	6	9k	1.63	N	4.93	N
Ply	8	8	8k	0.17	N	3.31	N
Podiatrist	6	6	17k	0.12	N	4.63	N
Rabies	4	4	11k	0.76	N	3.83	Y
Rush	4	4	2k	14.68	N	2.90	N
Shot	9	11	1k	50.50	N	4.44	N
To sniff	14	14	5k	1.75	V	4.17	Y
To spare	13	13	2k	14.14	V	3.48	N
Spasm	3	3	8k	0.54	N	4.07	Y
To spit	3	4	4k	8.85	V	4.71	N
To squint	7	7	6k	0.52	V	4.30	N
To step off*	11	11	1k	39.85	V	4.54	N
Stitch	3	3	4k	1.76	N	4.32	N
To suck	8	9	2k	13.97	V	3.55	N
To sweep	8	9	2k	4.54	V	3.72	N
To wean	6	6	8k	0.13	V	2.59	N
To withdraw	2	2	3k	2.77	V	3.04	N

Note. * In the case of compound nouns and phrasal verbs, only the SUBTLEX$_{US}$ percentage and the concreteness mean of the lexical verb (for phrasal verbs) or the less frequent word (for compound nouns) is reported.

APPENDIX F - SAMPLE MATERIALS

Please note: for the rest of the materials not included in Appendix F (i.e., vocabulary tests and tasks, while-watching tasks, and comprehension tests), please check the CD-ROM attached to the thesis.

F.1. - Vocabulary pre-, post- and delayed post-test

Parte I

A continuación escucharás veinte palabras en inglés. Escríbelas en inglés y tradúcelas al castellano o catalán. Si de alguna palabra conoces más de un significado, escríbelo. Escucharás cada palabra un total de dos veces.

	Inglés	Castellano - Catalán
1		
2		
3		
4		
5		
6		
7		
8		
9		
10		
11		
12		
13		
14		
15		
16		
17		
18		
19		
20		

Note. The test above only contains the first 20 TWs, since the rest (21-35 / 21-40) were in Part II. All the vocabulary tests used throughout the intervention followed exactly the same format.

F.2. - Vocabulary pre-tasks

F.2.1. - Grade 6 (G6E17).

<u>"Crazy 10-Minute Sale"</u>

1. Match each image with the most appropriate word from the box:

Sneakers / Spell / Commitment / Drill / Wand

F.2.2. - Grade 10 (G10E09)**.**

"The Trip – Part I"

1. Fill in the blanks with the appropriate words; the first letter is already given for you. Use the definitions to help you.

A) If you ask me, I think she is a complete n_______________; you never know how she will react.
B) I told my mum I wanted a ring for my birthday, but she bought me an a_______________, which I totally dislike.
C) Some hackers leaked Steven Spielberg's t_______________ before the movie was released.
D) The baby was covered with a b_______________ because it was really cold that night.
E) It's good to have c_______________ b_______________ in the airports! They are much more comfortable than carrying your suitcases with you.

Definitions

A) Strange or crazy person.
B) Piece of jewellery worn around the ankle.
C) Preliminary outline of a film specifying the key scenes, characters, and settings.
D) Large cover, often made of wool, used especially on beds to keep people warm.
E) A continuous moving band used for transporting goods from one part of a building to another.

F.3. - While-watching tasks

F.3.1. - Grade 6 (G6E17).

<u>"Crazy 10-Minute Sale"</u>

1. Escoge la respuesta correcta (A, B o C).

¿Qué días tiene clase de magia Alex?

A) Lunes y miércoles.
B) Miércoles y viernes.
C) Martes y jueves.

¿Cuántos días consecutivos ha aguantado Alex haciendo pucheros a sus padres?

A) 4 días.
B) 10 días.
C) 14 días.

F.3.2. - Grade 10 (G10E09).

<u>"The Trip – Part I"</u>

1. Escoge la respuesta correcta (A, B o C).

¿Cuál fue la última película en la que participó Helene?

A) Un corto de Three Stooges.
B) Un corto de los hermanos Marx.
C) Un corto de Charles Chaplin.

¿Con qué confunde el guarda de seguridad la crema hidratante de George?

A) Con loción de afeitar.
B) Con una maquinilla de afeitar.
C) Con un cuchillo.

F.4. - Vocabulary post-tasks

F.4.1. - Grade 6 (G6E17).

<u>"Crazy 10-Minute Sale"</u>

1. Escucharás cinco palabras en inglés; cada palabra se va a repetir dos veces. Escribe, en inglés, cada una de las palabras en el espacio en blanco y di qué significan (opción a, b, c…). Si no sabes qué quiere decir alguna palabra, elige la opción (f) 'No lo sé'.

1) ____________________

a) Botas
b) Zapatillas
c) Altavoz
d) Ladrar
e) Sollozar
f) No lo sé

2) ____________________

a) Comprar
b) Oler
c) Hechizo
d) Bruja
e) Cartero
f) No lo sé

3) ____________________

a) Esconder
b) Siesta
c) Cena familiar
d) Compromiso
e) Tratamiento
f) No lo sé

4) ____________________

a) Matar
b) Taladro
c) Destornillador
d) Guardería
e) Global
f) No lo sé

5) ____________________

a) Desfigurar
b) Bastón
c) Varita
d) Rubio
e) Imitación
f) No lo sé

F.4.2. - Grade 10 (G10E09).

"The Trip – Part I"

1. Escucharás cinco palabras en inglés; cada palabra se va a repetir dos veces. Escribe, en inglés, cada una de las palabras en el espacio en blanco y di qué significan (opción a, b, c…). Si no sabes qué quiere decir alguna palabra, elige la opción (f) 'No lo sé'.

1) ____________________

a) Diligente
b) Pasarela
c) Collar de perlas
d) Bala
e) Pulsera tobillera
f) No lo sé

2) ____________________

a) Banquete
b) Manta
c) Almohada
d) Crema hidratante
e) Inscribirse
f) No lo sé

3) ____________________

a) Guion
b) Suplemento
c) Medicación
d) Arropar
e) Loco
f) No lo sé

4) ____________________

a) Cinturón de seguridad
b) Cinta transportadora
c) Escalera mecánica
d) Vestíbulo
e) Excluir
f) No lo sé

5) ____________________

a) Extraño
b) Chiflado
c) Refugio
d) Cómico
e) Amablemente
f) No lo sé

F.5. - Comprehension tests

F.5.1. - Grade 6 (G6E17).

"Crazy 10-Minute Sale"

1. Marca si estos enunciados son verdaderos (V) o falsos (F).

V / F Gigi y Alex son enemigas desde pequeñas.
V / F Las amigas de Gigi llevan tiritas en la nariz porque se han dado un golpe.
V / F Max cubrirá a Alex mientras esté en las rebajas a cambio de un regalo.
V / F La nueva varita de Max también es un reproductor de vídeo.
V / F Finalmente, el padre de Alex la castiga dos meses sin postre.

2. Escoge la respuesta correcta (A, B o C).

¿Por qué ha salido mal el hechizo que ha hecho Justin?
A) Porque no se ha concentrado lo suficiente.
B) Porque no ha pronunciado las palabras correctas.
C) Porque aún no está preparado para hacer este tipo de hechizos.

¿Por qué Alex quiere duplicarse?
A) Porque así podrá practicar el nuevo hechizo que ha aprendido.
B) Porque así podrá ir a las rebajas.
C) Porque Justin la ha retado a hacerlo.

¿Cómo ha sido capaz Alex de reconocer a Harper entre tanta gente?
A) Porque sabía que estaba en primera fila.
B) Porque han quedado que se encontrarían en un lugar determinado.
C) Porque sabía cómo iba vestida.

¿Por qué Gigi ha traído a una mujer mayor a las rebajas?
A) Porque así serán dos y comprarán más cosas.
B) Porque la ayudará a elegir una chaqueta para su abuela.
C) Porque le bloqueará el paso a Alex.

¿Por qué Alex se mueve como quiere Max?
A) Porque la nueva varita de Max así se lo permite.
B) Porque Alex hizo mal el hechizo para duplicarse.
C) Porque Justin ha hecho un nuevo hechizo.

3. Ordena cronológicamente (de 1 a 8) estos hechos que pasan en el capítulo que acabas de ver. El '1' y el '5' te pueden servir de guía para ordenar los demás.

________ Jerry da dinero a 'Alex' para que se compre alguna cosa en las rebajas.
1 El episodio empieza con Justin duplicando un conejo.
________ Alex encuentra la chaqueta que tanto deseaba.
________ Harper se encuentra con la madre de Alex en las rebajas.
________ Jerry, el padre de Alex, no la deja ir de rebajas.
5 La tienda abre sus puertas para las rebajas.
________ Jerry se entera de que Justin le ha roto su taladro nuevo.
________ Descubrimos que Max quiere una nueva varita mágica.

F.5.2. - Grade 10 (G10E09).

"The Trip – Part I"

1. Marca si estos enunciados son verdaderos (V) o falsos (F).

V / F Kramer dice que se basa en los colores para aprenderse sus guiones.
V / F Helene ha pasado más de 60 años sin trabajar.
V / F A George le dan miedo los detectores de metales porque siempre pitan.
V / F Al final, tanto George como Jerry no quieren que Lupe meta las sábanas debajo de la cama.
V / F George cree que ha mantenido una conversación seria y decente con Corbin Bernsen y George Wendt.

2. Escoge la respuesta correcta (A, B o C).

¿Con quién estaba prometida Helene?
A) Con Mr. Sugarman.
B) Con Mickey Rooney.
C) Con Murphy Brown.

¿Por qué George lleva tanta ropa a Los Ángeles?
A) Porque no podía decidir que ropa llevar y se la llevó toda.
B) Porque no sabe qué tiempo hará en Los Ángeles.
C) Porque no sabe de qué humor estará.

¿Por qué Kramer quiere hablar con Fred Savage?
A) Porque quiere darle una copia del guión.
B) Porque es fan suyo y quiere darle una carta.
C) Porque quiere hacerse una foto con él.

¿Por qué Lupe no debería haber tirado el papelito de Jerry?
A) Porque había escrito unos chistes que pensaba explicar en el show.
B) Porque tenía apuntada la dirección del estudio.
C) Porque tenía apuntado el teléfono de Kramer.

¿Por qué la policía cree que Kramer es el principal sospechoso del asesinato?
A) Porque han encontrado su guión junto al cadáver de la joven.
B) Porque un testigo lo ha delatado.
C) Porque recogieron una foto suya en el lugar del crimen.

3. Ordena cronológicamente (de 1 a 8) estos hechos que pasan en el capítulo que acabas de ver. El '1' y el '5' te pueden servir de guía para ordenar los demás.

_______ Los policías encuentran el cadáver de una chica de 21 años.
_______ Lupe dice que tiró el papelito de Jerry cuando limpió la habitación.
1 El episodio empieza con Jerry y George hablando sobre Kramer y su participación en un programa de televisión.
_______ Descubrimos que George usa crema hidratante.
_______ Jerry anuncia que tiene tickets gratis para ir a Los Ángeles.
5 Kramer se encuentra con Fred Savage.
_______ Jerry actúa en *The Tonight Show*.
_______ George le cuenta la historia del gato a Corbin Bernsen.

APPENDIX G - PAUSING TIMES FOR THE WHILE-WATCHING TASKS

Table G.1
Pausing times for the while-watching tasks in Grades 6 and 10 and at university.

Grade 6				Grade 10 and university			
Episode	**Time**	**Episode**	**Time**	**Episode**	**Time**	**Episode**	**Time**
G6E01	10m 38s	**G6E13**	10m 56s	**G10E01**	10m 59s	**G10E13**	11m 08s
G6E02	10m 26s	**G6E14**	10m 15s	**G10E02**	12m 19s	**G10E14**	11m 06s
G6E03	09m 10s	**G6E15**	10m 38s	**G10E03**	11m 05s	**G10E15**	09m 44s
G6E04	09m 56s	**G6E16**	12m 18s	**G10E04**	12m 27s	**G10E16**	12m 04s
G6E05	12m 04s	**G6E17**	11m 04s	**G10E05**	13m 45s	**G10E17**	10m 45s
G6E06	10m 42s	**G6E18**	11m 59s	**G10E06**	12m 52s	**G10E18**	11m 34s
G6E07	12m 42s	**G6E19**	11m 03s	**G10E07**	12m 19s	**G10E19**	09m 40s
G6E08	10m 15s	**G6E20**	10m 41s	**G10E08**	13m 41s	**G10E20**	11m 00s
G6E09	10m 36s	**G6E21**	11m 27s	**G10E09**	08m 25s	**G10E21**	11m 03s
G6E10	13m 10s	**G6E22**	12m 10s	**G10E10**	10m 08s	**G10E22**	11m 33s
G6E11	10m 37s	**G6E23**	11m 06s	**G10E11**	11m 01s		
G6E12	12m 08s	**G6E24**	11m 18s	**G10E12**	11m 00s		

APPENDIX H - PROFICIENCY TESTS

H.1. - Oxford Placement Test (example)

Page 1

Oxford Placement Test 1

Grammar Test PART 1

Name	....
Total Listening	.... / 100
Total Grammar	.... / 100
Grand Total	.... **/ 200**

Look at these examples. The correct answer is ticked.

a In warm climates people | like ✓ | likes | are liking | sitting outside in the sun.

b If it is very hot, they sit | at | in ✓ | under | the shade.

Now the test will begin. Tick the correct answers.

1 Water | is to boil | is boiling | boils | at a temperature of 100°C. 1 ______

2 In some countries | there is | is | it is | very hot all the time. 2 ______

3 In cold countries people wear thick clothes | for keeping | to keep | for to keep | warm. 3 ______

4 In England people are always talking about | a weather | the weather | weather |. 4 ______

5 In some places | it rains | there rains | it raining | almost every day. 5 ______

6 In deserts there isn't | the | some | any | grass. 6 ______

7 Places near the Equator have | a warm | the warm | warm | weather even in the cold season. 7 ______

8 In England | coldest | the coldest | colder | time of year is usually from December to February. 8 ______

9 | The most | Most of | Most | people don't know what it's really like in other countries. 9 ______

10 Very | less | little | few | people can travel abroad. 10 ______

11 Mohammed Ali | has won | won | is winning | his first world title fight in 1960. 11 ______

12 After he | had won | have won | was winning | an Olympic gold medal he became a professional boxer. 12 ______

13 His religious beliefs | have made him | made him to | made him | change his name when he became champion. 13 ______

14 If he | has | would have | had | lost his first fight with Sonny Liston, no one would have been surprised. 14 ______

15 He has travelled a lot | both | and | or | as a boxer and as a world-famous personality. 15 ______

subtotal /15

Oxford Placement Test 1
Listening Test

Name	....
Total Listening	.... / 100
Total Grammar	.... / 100
Grand Total	.... **/ 200**

Look at the example below. Listen to the tape. You will hear the example *once* only. Decide which word you hear, 'soap', or 'soup'.

a Will you get me some | soap | soup ✓ | at the supermarket?

The word was 'soup', so 'soup' is ticked. Now look at these examples, and listen to the tape again. This time, you tick the words you hear. For example, if you hear 'shorts', tick 'shorts'.

b The team need new | shirts | shorts | .

c They've recently developed a new kind of | vine | wine | around here.

The words on the tape were 'shorts' and 'vine', so the correct answers look like this:

b The team need new | shirts | shorts ✓ | .

c They've recently developed a new kind of | vine ✓ | wine | around here.

Now the test will begin. Listen to the tape and tick (✓) the words you hear.

1 I gather you've been having trouble with your | earring | hearing | . 1 ______

2 A number of students are expected to join the advanced | composition | conversation | class. 2 ______

3 This beard of mine is awfully itchy. I'll be glad when it | goes | grows | . 3 ______

4 I doubt if he's very comfortable in his | present | prison | bed. 4 ______

5 Have you played | Dennis | tennis | very much recently? 5 ______

6 Martina lives in a great big | freezing | Friesian | barn. 6 ______

7 Do you have any idea how long ago it was | found | founded | ? 7 ______

8 Your letter must have crossed with | my own | mine | . 8 ______

9 One thing I really | loved | loathed | in the late nineties was the style of the clothes. 9 ______

10 My sister says | he's | she's | a very nice person. 10 ______

11 That Dutch friend of mine you met yesterday is a very good | chess | jazz | player. 11 ______

12 That's the Euro equivalent of | 30p | 40p | . 12 ______

13 Do we need to change the | cloths | clocks | tonight? 13 ______

14 Today's a | holiday | horrid day | , isn't it? 14 ______

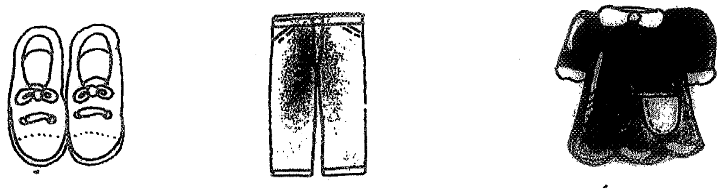
15 Well, I wonder what | joys | choice | they have in store for us this time. 15 ______

H.2. - **BAF listening test** (example)

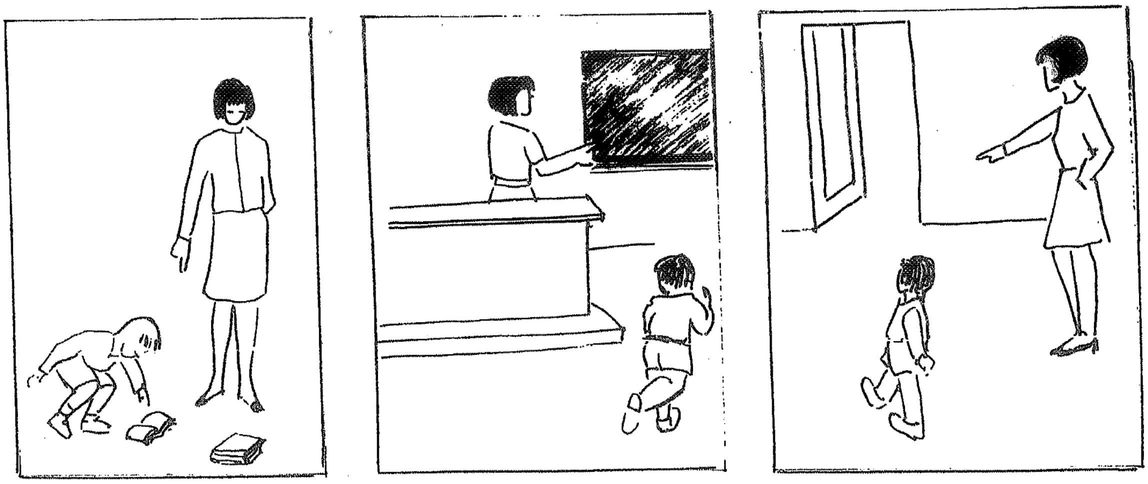
NOM: .... ESCOLA: ....

CURS: .... DATA: ....

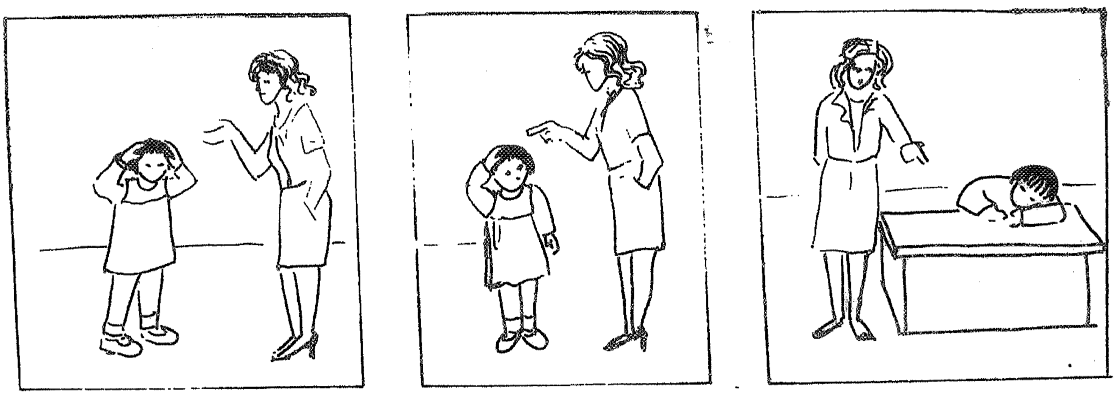
1.



2.



3.



APPENDIX I - SAMPLE MATERIALS OF THE PILOT STUDY

Please note: for the rest of the materials not included in Appendix I (i.e., vocabulary tests and tasks, while-watching tasks, comprehension tests, and questionnaires), please check the CD-ROM attached to the thesis.

I.1. - Vocabulary pre-test

A continuación escucharás ocho palabras en inglés. Escríbelas en inglés y tradúcelas al castellano o catalán. Escucharás cada palabra un total de tres veces.

	Inglés	**Castellano - Catalán**
1		
2		
3		
4		
5		
6		
7		
8		

I.2. - Vocabulary pre-tasks

I.2.1. - *Boy Meets World.*

"On the Fence"

1. In pairs, answer the following questions:

- Do you help with the housework? If you do, do you do it voluntarily? If you don't, why not?
- Do you ever get pocket money from any members of your family? If you do, what do you do with it? If you don't, what would you do with it?
- Have you ever helped at home in exchange of some money?

2. Fill in the gaps with the following words:

Bucks / Water gun / War / Kid / Fence / Shutters / Remember / Stuff

- You should tidy up your room. Put all your __________________ away!
- __________________ destroys whole countries. I hope the world will be in peace someday.
- If we go to the swimming pool, we could play with our __________________!
- I need 10 __________________ to buy myself a ticket for the cinema.
- You should always __________________ to switch the lights off when you leave the room.
- When you are a _________________, you don't have to work. School is your only obligation.
- Can you please close the __________________ so the light doesn't go through the window?
- We put a __________________ around our garden so that the dog couldn't escape.

I.2.2. - *Seinfeld.*

"The Movie"

1. In pairs, answer the following questions:

- How often do you go to the cinema?
- Do you usually buy your movie tickets in advance?
- Have you ever been to the cinema on your own? If so, when? If not, why not?
- Which row do you prefer to sit on when you go to the cinema?

2. Fill in the gaps with the following words:

Coming attractions / Spot / Owe / Break / Rip / Stub / Nostrils / Terrific

- Remember last week you lent me 10€. I'm so forgetful! I still __________________ you that money!
- Can you __________________ a 5€-note into coins? The vending machine doesn't accept notes!
- We should go there in advance to get a good __________________ and not miss a thing!
- The little hairs showing out of his __________________ are a bit disgusting.
- You should always keep your __________________ with you in case you want to go out and back into the theater later.
- It was a __________________ movie. We had a fantastic time!
- Be careful when you __________________ the ticket in half; make sure the barcode isn't damaged.
- I hate going to this cinema! They show endless __________________ before the actual movie starts!

I.3. - While-watching task

"On the Fence / The Movie"

1. Contesta la siguiente pregunta:

¿Qué ha pasado hasta el momento en este capítulo?

I.4. - Vocabulary post-tasks

I.4.1. - *Boy Meets World*.

"On the Fence"

1. Escucharás ocho palabras en inglés. Cada palabra se va a repetir tres veces. Di qué significan estas palabras (opción a, b, c…). Si no sabes qué quiere decir alguna palabra, elige la opción (f) 'No lo sé'.

1) **Remember**

a) Reflexionar
b) Recordar
c) Miembro
d) Visitar
e) Hielo
f) No lo sé

2) **Stuff**

a) Globo
b) Cerca
c) Trastos
d) Trabajadores
e) Cosas
f) No lo sé

3) **Fence**

a) Reducir
b) Precio
c) Obstáculo
d) Valla
e) Bailar
f) No lo sé

4) **Bucks**

a) Dólar
b) Cupón
c) Espalda
d) Rosas
e) Inteligente
f) No lo sé

5) **Shutters**

a) Inaugurar
b) Mantequilla
c) Persianas
d) Carpa
e) Fontanero
f) No lo sé

6) **Water gun**

a) Baño
b) Llorar
c) Manguera
d) Pistola de agua
e) Diversión
f) No lo sé

7) **War**

a) Guerra
b) Puerta
c) Torneo
d) Juguete
e) Vender
f) No lo sé

8) **Kid**

a) Adolescente
b) Niño/a
c) Cerdo
d) Helado
e) Gritar
f) No lo sé

I.4.2. - *Seinfeld.*

"The Movie"

1. Escucharás ocho palabras en inglés. Cada palabra se va a repetir tres veces. Escribe estas palabras por orden en el espacio libre (1, 2, 3…) y di qué significan (opción a, b, c…). Si no sabes qué quiere decir alguna palabra, elige la opción (f) 'No lo sé'.

1) ____________________

a) Abrigo
b) Extraño
c) Arco
d) Deber
e) Obligar
f) No lo sé

2) ____________________

a) Ajetreo
b) Brillar
c) Ombligo
d) Orificios nasales
e) Anginas
f) No lo sé

3) ____________________

a) Ranura
b) Lugar
c) Terreno
d) Persona repelente
e) Ascender
f) No lo sé

4) ____________________

a) Zapatero
b) Pensativo
c) Certificado
d) Resguardo
e) Puñalada
f) No lo sé

5) ____________________

a) Bulto
b) Sellar
c) Rasgar
d) Propina
e) Contratar
f) No lo sé

6) ____________________

a) Embarcadero
b) Terrorífico
c) Genial
d) Talentoso
e) Chanchullo
f) No lo sé

7) ____________________

a) Modificar
b) Dar cambio
c) Frenar
d) Viaje organizado
e) Muy grande
f) No lo sé

8) ____________________

a) Créditos de una película
b) Gabardina
c) Mezclar
d) Atracciones de feria
e) Tráiler
f) No lo sé

I.5. - Comprehension tests

I.5.1. - *Boy Meets World.*

"On the Fence"

1. Marca si estos enunciados son verdaderos (V) o falsos (F).

V / F Cory tiene envidia de su hermano porque se puede comprar las cosas que quiere.
V / F Cory quiere trabajar como reponedor en el mercado.
V / F El Sr. Feeny tiene que corregir muchas redacciones.
V / F El Sr. Feeny no quiere pagar a Cory por avanzado porque no tiene dinero en efectivo.
V / F Cory ha comprado una pistola de agua pequeña para su madre.

2. Escoge la respuesta correcta (A, B o C).

¿Qué súper héroe escogería Cory para que fuese su padre?
A. Batman.
B. Capitán América.
C. Superman.

¿Por qué Cory dice que los globos de agua son mejores que las pistolas?
A. Porque los globos de agua mojan más que las pistolas.
B. Porque él no tiene una pistola de agua.
C. Porque la pistola de agua se tiene que recargar muy a menudo.

¿Por qué Cory tira piropos a su madre?
A. Porque quiere que le compre algo.
B. Porque lleva un vestido muy elegante.
C. Porque quiere que le haga su comida favorita.

¿Por qué Cory no puede ir a la guerra de agua?
A. Porque está resfriado.
B. Porque sus amigos no le han invitado.
C. Porque tiene que pintar las persianas.

¿Cómo ha conseguido Cory las pistolas de agua pequeñas?
A. Sus amigos se las han regalado.
B. Las ha comprado con el dinero que le sobraba.
C. Ha cambiado su pistola grande por las pequeñas.

3. Ordena cronológicamente (de 1 a 8) estos hechos que pasan en el capítulo que acabas de ver. El '1' y el '5' te pueden servir de guía para ordenar los demás.

__1__ El episodio empieza con Cory y sus amigos hablando de súper héroes.
_______ El Sr. Feeny confirma que las matemáticas no son el fuerte de Cory.
_______ Cory tiene que pedir dinero a su madre para comprar más pintura.
__5__ Cory pinta la valla del jardín por error.
_______ Los padres de Cory le dicen que le regalarán la pistola por Navidad.
_______ El padre de Cory acaba de pintar la valla.
_______ Cory decide buscarse un empleo.
_______ La hermana pequeña llama al servicio de emergencias.

I.5.2. - *Seinfeld.*

<u>"The Movie"</u>

1. Marca si estos enunciados son verdaderos (V) o falsos (F).

V / F Jerry había quedado con sus amigos para ver *Checkmate* a las 21:30h.
V / F El compañero de Jerry en el coche no come pescado porque cuando era pequeño le dio dolor de barriga.
V / F Kramer describe a George como un hombre cabezón y unos orificios nasales muy abiertos.
V / F En 25 años, nadie ha pedido a George el resguardo de la entrada del cine.
V / F Elaine tendría que sentarse en primera fila para estar al lado de George.

2. Escoge la respuesta correcta (A, B o C).

¿Dónde había estado esperando Kramer?
A. En el Paragon.
B. En el Paradise Twin.
C. En el Papaya King.

¿Por qué no avanza la cola del cine?
A. Porque todos los de la cola tienen que comprar entrada.
B. Porque todas las entradas están vendidas.
C. Porque todos los de la cola ya tienen entrada.

¿Por qué George se muestra irónico con el señor que tiene delante en la cola?
A. Porque antes el señor no había estado preciso con la respuesta a su pregunta.
B. Porque antes el señor no le había contestado su pregunta.
C. Porque no le gusta hacer cola y siempre reacciona así.

¿Por qué Elaine está a punto de perder los nervios?
A. Porque ella no quiere comprar caramelos y George sí.
B. Porque a George no le gusta ninguna de sus propuestas.
C. Porque los chicos quieren ir a ver *Rochelle Rochelle*.

¿Con qué compara George la cara de Elaine?
A. Con un ángel.
B. Con una sartén.
C. Con una reina.

3. Ordena cronológicamente (de 1 a 8) estos hechos que pasan en el capítulo que acabas de ver. El '1' y el '5' te pueden servir de guía para ordenar los demás.

___**5**___ Elaine va diciendo que sus amigos están comprando palomitas porque así nadie ocupe sus asientos.
_______ Cuando Elaine vuelve a la sala, no sabe donde está su butaca.
_______ Los protagonistas deciden ir a otro cine porque no quedan entradas.
___**1**___ El episodio empieza con Jerry quejándose que el show va con retraso.
_______ Elaine se da cuenta que Kramer le quitó el sitio porque tiene su abrigo.
_______ Kramer quiere comerse un frankfurt de un restaurante y no del cine.
_______ Jerry y su compañero deciden compartir un taxi.
_______ El empleado del cine no deja entrar a George porque no tiene el resguardo de la entrada.

I.6. - Questionnaire

I.6.1. - *Boy Meets World*.

<u>"On the Fence"</u>

Nos gustaría saber tu opinión acerca del capítulo que acabas de ver. Por favor, marca con una cruz la opción más apropiada en cada pregunta:

	☺☺ (Mucho)	☺ (Bastante)	😐 (Ni mucho ni poco)	☹ (Poco)	☹☹ (Nada)
1. ¿Te ha gustado el capítulo que has visto?					
2. ¿Te gustaría ver más capítulos de esta serie?					
3. ¿Has entendido bien lo que ocurre en este capítulo?					
4. ¿Has leído los subtítulos?					
5. ¿Los subtítulos te han ayudado a entender el capítulo?					
6. ¿Crees que has aprendido algo de vocabulario en inglés?					

Otros comentarios:

____

____

I.6.2. - *Seinfeld.*

<u>"The Movie"</u>

Nos gustaría saber tu opinión acerca del capítulo que acabas de ver. Por favor, marca con una cruz la opción más apropiada en cada pregunta:

	Mucho	**Bastante**	**Ni mucho ni poco**	**Poco**	**Nada**
1. ¿Te ha gustado el capítulo que has visto?					
2. ¿Te gustaría ver más capítulos de esta serie?					
3. ¿Has entendido bien lo que ocurre en este capítulo?					
4. ¿Has leído los subtítulos?					
5. ¿Los subtítulos te han ayudado a entender el capítulo?					
6. ¿Crees que has aprendido algo de vocabulario en inglés?					

Otros comentarios:

____

____

APPENDIX J - DETAILED, GENERAL AND INFERENTIAL COMPREHENSION

As explained in the Methodology (see section 6.5.4.), the comprehension tests included questions tapping into three types of comprehension (detailed, general, and inferential) (Buck, 2001; Rodgers, 2013). Table J.1 shows the percentage of each question type according to the proficiency groups. It must be borne in mind that this classification just applies to the first two exercises (T/F and MC), excluding the last exercise in which learners had to order statements in chronological order.

Table J.1
Percentage of detailed, general and inferential questions according to proficiency group.

Question type (in %)	Grade 6	Grade 10	University
Detailed	37.92	38.45	52.50
General	36.25	43.04	36.25
Inferential	25.83	18.51	11.25

J.1. Results

The percentage of correct responses for each question type was calculated to explore whether detailed, general and inferential comprehension were differently enhanced by the viewing practice (see Table J.2 and Figure J.1 for the results).

Table J.2
Percentage of correct answers according to question type and proficiency group.

	Correct answers (in percentage)								
	University			**Grade 10**			**Grade 6**		
	Detailed	**General**	**Inferential**	**Detailed**	**General**	**Inferential**	**Detailed**	**General**	**Inferential**
Mean	76	81.85	87.13	66.15	74.54	74.64	64.94	66.61	62.90
SD	10.57	6.74	11.99	10.63	8.97	11.97	12.65	9.89	13.17
Min.	45.24	62.07	55.56	47.67	58.51	47.50	39.56	49.43	41.94
Max.	92.86	93.10	100	86.05	88.30	92.50	90.11	83.91	88.71
Sig. differences	✓			✓			✗		

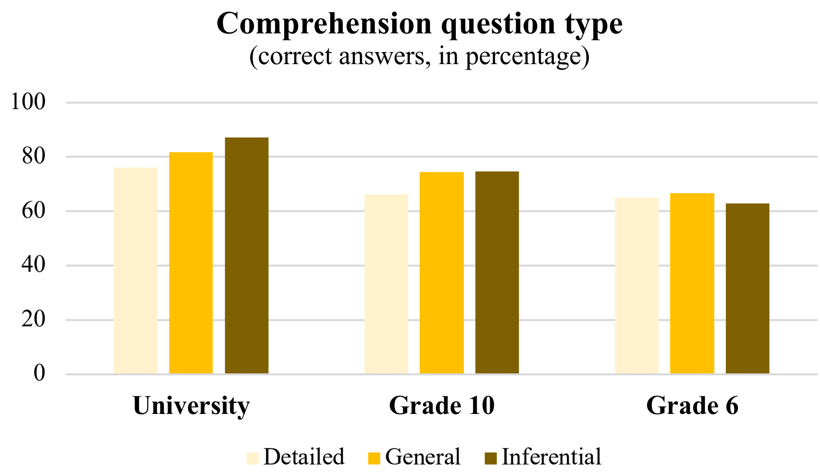


Figure J.1 - Percentage of correct answers according to question type and proficiency group.

J.1.1. University.

At university, Table J.2 shows that there were some differences in the percentage of correct responses given to the three question types: learners correctly answered inferential questions more often than general and detailed questions. A Friedman test (*N*=37), due to the non-normal distribution of the data, showed that there were significant differences in the percentage of correct responses depending on the type of question ($\chi^2(2) = 18.368$, $p = .000$), and pairwise comparisons further revealed that detailed questions were significantly more challenging to answer than general ($p = .025$) and inferential ($p = .000$) questions. No differences were observed between general and inferential questions.

J.1.2. Grade 10.

Table J.2 also shows some minor differences between the percentage of correct answers to general and inferential questions in Grade 10 learners. However, detailed questions were correctly answered less often than the other two types. Another Friedman test (*N*=23) showed that participants' responses significantly differed depending on the type of question they were asked to answer ($\chi^2(2) = 30.929$, $p = .000$). Pairwise comparisons further proved that detailed questions were more difficult to answer than general ($p = .000$) and inferential questions ($p = .000$). Results obtained for general and inferential questions did not significantly differ.

J.1.3. Grade 6.

Finally, in Grade 6, differences were minimal across question type (as seen in Table J.2 as well). This was later confirmed by a RM ANOVA (*N*=18), which revealed that there were no significant differences in the percentage of correct responses given to detailed, general and inferential questions ($F(2, 16) = 2.459$, $p = .111$, partial eta squared = .197).

J.2. Some remarks

Results showed that Grade 6 learners answered in a similar way all questions types, while, in Grade 10 and university, detailed questions were more challenging to answer than general and inferential questions. This indicates that viewing L1 subtitled TV series for an academic year did not promote detailed, general or inferential comprehension in beginner learners. However, general and inferential comprehension were significantly enhanced by viewing L2 subtitled TV series among high-school and university participants.

It makes sense that detailed questions were harder to answer than general and inferential questions for more proficient learners. In the less proficient group (sixth graders), the content comprehension achieved was limited, and these learners did not get very high scores on the tests, so their correct answer rate was poor, and no differences could be spotted depending on the type of question. In contrast, in more proficient groups, their comprehension was better, but not good enough to understand minor details of the episode. Grade 10 and university learners understood the general storyline of the episode and were also able to make inferences and connections between events, but likely failed to understand and retain secondary information not so relevant to the plot. It is also possible that, if they had had a higher proficiency level, they would have got more detailed questions right and differences would have disappeared. It should be pointed out, though, that more research comparing viewing subtitled and non-subtitled videos to listening alone would be needed to draw clearer conclusions on the type of comprehension being fostered by multimodal input.