



UNIVERSITAT DE  
BARCELONA

# Language learning through extensive TV viewing

## A study with adolescent EFL learners

Geòrgia Pujadas Jorba

**ADVERTIMENT.** La consulta d'aquesta tesi queda condicionada a l'acceptació de les següents condicions d'ús: La difusió d'aquesta tesi per mitjà del servei TDX ([www.tdx.cat](http://www.tdx.cat)) i a través del Dipòsit Digital de la UB ([diposit.ub.edu](http://diposit.ub.edu)) ha estat autoritzada pels titulars dels drets de propietat intel·lectual únicament per a usos privats emmarcats en activitats d'investigació i docència. No s'autoritza la seva reproducció amb finalitats de lucre ni la seva difusió i posada a disposició des d'un lloc aliè al servei TDX ni al Dipòsit Digital de la UB. No s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX o al Dipòsit Digital de la UB (framing). Aquesta reserva de drets afecta tant al resum de presentació de la tesi com als seus continguts. En la utilització o cita de parts de la tesi és obligat indicar el nom de la persona autora.

**ADVERTENCIA.** La consulta de esta tesis queda condicionada a la aceptación de las siguientes condiciones de uso: La difusión de esta tesis por medio del servicio TDR ([www.tdx.cat](http://www.tdx.cat)) y a través del Repositorio Digital de la UB ([diposit.ub.edu](http://diposit.ub.edu)) ha sido autorizada por los titulares de los derechos de propiedad intelectual únicamente para usos privados enmarcados en actividades de investigación y docencia. No se autoriza su reproducción con finalidades de lucro ni su difusión y puesta a disposición desde un sitio ajeno al servicio TDR o al Repositorio Digital de la UB. No se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR o al Repositorio Digital de la UB (framing). Esta reserva de derechos afecta tanto al resumen de presentación de la tesis como a sus contenidos. En la utilización o cita de partes de la tesis es obligado indicar el nombre de la persona autora.

**WARNING.** On having consulted this thesis you're accepting the following use conditions: Spreading this thesis by the TDX ([www.tdx.cat](http://www.tdx.cat)) service and by the UB Digital Repository ([diposit.ub.edu](http://diposit.ub.edu)) has been authorized by the titular of the intellectual property rights only for private uses placed in investigation and teaching activities. Reproduction with lucrative aims is not authorized nor its spreading and availability from a site foreign to the TDX service or to the UB Digital Repository. Introducing its content in a window or frame foreign to the TDX service or to the UB Digital Repository is not authorized (framing). Those rights affect to the presentation summary of the thesis as well as to its contents. In the using or citation of parts of the thesis it's obliged to indicate the name of the author.

**Doctoral dissertation**

Doctoral program in Linguistic, Literary and Cultural Studies

**LANGUAGE LEARNING  
THROUGH EXTENSIVE TV VIEWING**  
A study with adolescent EFL learners

by

Geòrgia Pujadas Jorba

Supervised by Dra. Carme Muñoz



UNIVERSITAT DE  
BARCELONA

**2019**



## Abstract

This doctoral dissertation explores the benefits of an extensive exposure to L2 television for content comprehension and vocabulary learning with beginner, adolescent EFL learners, through a longitudinal classroom-based intervention. The study also aims at investigating the role of frequency and imagery in word learning, and evaluating the long-term effects this type of intervention may have on learners' attitudes towards the use of L2 television for language learning purposes. In order to do so, several aspects were examined, including factors related to the intervention, the learner, and the input itself. Four classes of Grade 8 students viewed 24 consecutive episodes of a TV series – spaced over an academic year – under four experimental conditions, with each class being assigned to a different treatment. Two classes viewed the episodes with English [L2] captions, and two with Spanish [L1] subtitles. One class in each language condition received, additionally, explicit instruction on target lexical items. The first study in this thesis looked at the differential effect of captions and subtitles on viewing comprehension, alongside several mediating factors. Results confirmed the higher efficiency of subtitles over captions for content comprehension at this level of proficiency, and the importance of prior vocabulary knowledge when viewing with captions. It was also found that the episodes' lexical coverage was a strong predictor of comprehension, although no pattern of improvement could be observed over time – even though learners' perceived comprehension increased by the end of the intervention. In the second study, word-form and word-meaning gains were examined, following a pre- / post-test design. Results revealed that having explicit instruction on vocabulary (i.e. being pre-taught the words through short pre-viewing activities) yielded significantly higher vocabulary gains, and that proficiency played a key role in how learners made use of this type of input.

Language of the on-screen text, however, did not emerge as a predictor of gains, although when captions were displayed this tended to lead to higher learning, especially when combined with instruction. Results also indicated that there was a positive correlation between vocabulary gains and comprehension, and that a high percentage of the vocabulary learnt was retained in the long term. The third study focused on the effects of word repetition, spacing, and imagery support. Analysis showed that words with a higher number of encounters or that appeared in a massed condition (i.e. in the same episode) were better learnt. It was also found that words that were image-supported had also higher learning gains. The last study looked into learners' perceptions and feeling of learning from viewing audio-visual input in English through questionnaires and interviews. Students reported being highly motivated to learn through L2 videos, finding them useful for a number of language aspects, including vocabulary learning, listening comprehension, and matching aural and written forms of the words. Data also showed a shift in viewing habits in the long term, indicating the appropriateness of this type of classroom intervention to foster autonomous viewing at home. Taken as a whole, the results from this dissertation provide evidence that extensive viewing of captioned and subtitled TV series supports comprehension and L2 vocabulary learning. The characteristics of this type of input (i.e. repeated encounters with words, imagery) have been shown to contribute to facilitate language learning. Additionally, EFL learners, at this age and proficiency level, are motivated to learn through this media – in and outside the formal setting.

## Resum

Aquesta tesi doctoral explora els beneficis d'una exposició prolongada a programes de televisió en llengua estrangera per a la comprensió del contingut i l'adquisició de vocabulari amb estudiants d'anglès novells i adolescents, a través d'una intervenció longitudinal a l'aula. L'estudi també té per objectiu investigar el rol de la freqüència de les paraules i la presència de la imatge en l'aprenentatge de vocabulari, així com avaluar els efectes a llarg termini que una intervenció d'aquestes característiques pot tenir en les actituds dels alumnes a l'hora d'utilitzar la televisió com a eina d'aprenentatge. Amb aquest propòsit, s'han examinat diversos aspectes relacionats amb la intervenció, els estudiants i els materials audiovisuals. Quatre classes de 2n d'ESO han vist 24 episodis consecutius d'una sèrie de televisió (al llarg d'un any acadèmic) en quatre condicions experimentals, i cada classe ha estat assignada a un tractament diferent. Dues classes han vist els episodis amb subtítols en anglès (L2), i dues amb subtítols en castellà (L1). Una classe de cada, a més a més, ha rebut formació explícita sobre el vocabulari dels episodis. El primer estudi de la tesi ha investigat de quina manera afecten els subtítols en anglès i castellà el nivell de comprensió dels episodis, juntament amb altres factors. Els resultats han confirmat que, de cara a la comprensió del contingut (a aquesta edat i competència lingüística), els subtítols en llengua nadiua són més eficaços, i que el nivell de vocabulari en anglès juga un paper significatiu quan la sèrie es mira amb subtítols en aquesta llengua. El nivell de complexitat lèxica dels episodis (*lexical coverage*) també ha emergit com a predictor significatiu del nivell de comprensió, però no s'ha pogut observar cap millora a mesura que els estudiants han anat veient més episodis (tot i que percebien que entenien millor la sèrie al final de la intervenció). En el segon estudi s'ha examinat l'aprenentatge de la forma i el significat de paraules noves, a través d'un

disseny de pre- i post-test. Els resultats han revelat que l'ensenyament explícit de vocabulari (a través d'activitats curtes prèvies a la visualització) comporta guanys significativament majors, i que el nivell de competència en la llengua estrangera juga un paper clau a l'hora de beneficiar-se d'aquest tipus d'input. Tot i que la llengua dels subtítols no ha predit els guanys en vocabulari, els grups amb accés a subtítols en anglès han après un percentatge una mica més elevat de paraules, especialment quan es combinava amb l'ensenyament explícit d'element lèxics. També s'ha trobat una correlació positiva entre els guanys en vocabulari i el nivell general de comprensió, i que un elevat percentatge de les paraules apreses es retenen a llarg termini. El tercer estudi s'ha centrat en els efectes de la freqüència, la distribució i el suport visual de les paraules. Les anàlisis han mostrat que les paraules que més es repeteixen, i les que apareixen concentrades en el mateix episodi, són més fàcils d'aprendre. Les paraules que apareixen acompanyades de la seva imatge també s'han après més. L'últim estudi de la tesi ha explorat les percepcions dels estudiants sobre l'ús de material audiovisual, així com la sensació d'aprenentatge que se'n deriva, a través de qüestionaris i entrevistes. Els estudiants han expressat una elevada motivació per a aprendre anglès a través d'aquest tipus de materials, i consideren que són útils per a diversos aspectes lingüístics, com ara l'aprenentatge de vocabulari, la comprensió oral i la possibilitat d'establir la connexió entre la forma oral i escrita de les paraules. Les dades també revelen un canvi d'hàbits de consum de material audiovisual a llarg termini, un fet que indica que una intervenció d'aquestes característiques a l'aula és apropiada per a fomentar que els alumnes continuïn veient sèries en anglès a casa. En conjunt, aquesta tesi doctoral demostra que una exposició prolongada a sèries de televisió amb subtítols en anglès o castellà facilita la comprensió i l'aprenentatge de vocabulari en anglès. Les característiques d'aquest tipus d'input (p. e. la repetició del vocabulari i la presència de les imatges) afavoreixen l'aprenentatge de la llengua estrangera. A més a més, els estudiants estan motivats per a aprendre a través d'aquest mitjà, dins i fora de l'aula.

## Acknowledgements

First and foremost, I would like to express my sincerest gratitude to my supervisor, Dr. Carme Muñoz, who inspired me to start this project and has encouraged me at every step along the way. Her feedback and comments during the whole process have contributed greatly to the outcomes of the study, and I have learnt a great deal from her.

I would also like to thank Dr. Elke Peters for her valuable advice and insightful comments during my research stay at KU Leuven, and Dr. Michael Rodgers for taking the time to sit down with me and discuss our mutual field of interest during my short visit to Carleton University.

I would like to extend my thanks to Dr. Imma Miralpeix and Dr. Robert Vanderplank - along with Dr. Peters – for being part of my thesis committee and for their helpful, expert advice in the earlier stages of my dissertation.

I would also like to thank the GRAL Research Group, as well as my friends and colleagues at the University of Barcelona, for lending me a hand during data collection, solving problems together or just sharing a coffee and encouraging me during the difficult moments.

I am indebted with the secondary school who accepted to take part in this study, and for allowing a young researcher to roam around the school for over a year. Special thanks to Dr. Dolors Ortega and Jordi Martínez, who were responsible for making it happen. Big thanks to the teachers and especially to the students, without whom this study would have not been possible (nor half as fun as it was). Special thanks to those students who made sure that the sessions were properly conducted and to those who agreed to be



interviewed. Thanks for making the data collection process one of the most rewarding parts of this thesis.

I am particularly thankful for the people who supported me during this endeavour. Special thanks to Marta and Sílvia for their unconditional moral support and thanks to Natasha and Eva for always being there, ready to go on an adventure or to engage in a deep discussion about our research projects.

I would also like to thank my parents, Jaume and M<sup>a</sup> Teresa and my brother Arnau, who supported and encouraged me at each step of the way, and who took care of me in times of struggle. I want to dedicate this thesis especially to my father, who instilled in me the craving for knowledge that has led me to where I am today.

Finally, I would like to thank my supervisor again, as thanking her once does not seem enough. Thank you, Carme, for your constant support.

# Table of contents

<b>Abstract</b> .....	<b>i</b>
<b>Resum</b> .....	<b>iii</b>
<b>Acknowledgements</b> .....	<b>v</b>
<b>Table of contents</b> .....	<b>vii</b>
<b>List of tables</b> .....	<b>xi</b>
<b>List of figures</b> .....	<b>xv</b>
<b>List of appendices</b> .....	<b>xix</b>
<b>Chapter 1. Introduction</b> .....	<b>1</b>
<b>Chapter 2. Literature review</b> .....	<b>5</b>
2.1. Introduction .....	5
2.2. Audio-visual input and second language learning .....	7
2.3. Captions and subtitles.....	19
2.4. Viewing comprehension.....	30
2.5. Vocabulary learning .....	46
2.6. Attitudes towards OV input .....	70
<b>Chapter 3. Rationale</b> .....	<b>81</b>
3.1. Introduction .....	81
3.2. Research questions .....	82
<b>Chapter 4. Methodology</b> .....	<b>85</b>
4.1. Introduction .....	85
4.2. Participants .....	86
4.3. Audio-visual materials.....	91
4.4. Testing instruments.....	103
4.5. Procedure .....	125

<b>Chapter 5. Viewing comprehension .....</b>	<b>131</b>
5.1. Introduction .....	131
5.2. Research questions .....	133
5.3. Participants .....	134
5.4. Materials and procedure .....	134
5.5. Preliminary analysis .....	135
5.6. Results .....	139
5.7. Summary of findings .....	154
5.8. Discussion.....	155
<b>Chapter 6. Vocabulary learning .....</b>	<b>163</b>
6.1. Introduction .....	163
6.2. Research questions .....	165
6.3. Part one: vocabulary learning .....	166
6.4. Part two: long term retention.....	184
6.5. Summary of findings .....	193
6.6. Discussion.....	194
<b>Chapter 7. The role of frequency and imagery .....</b>	<b>201</b>
7.1. Introduction .....	201
7.2. Research questions .....	203
7.3. Part one: the role of frequency.....	204
7.4. Part two: imagery analysis .....	214
7.5. Summary of findings .....	224
7.6. Discussion.....	225
<b>Chapter 8. The learners' perspective .....</b>	<b>229</b>
8.1. Introduction .....	229
8.2. Research questions .....	230
8.3. Participants .....	231
8.4. Materials and procedure .....	232
8.5. Preliminary analysis .....	234
8.6. Preferences.....	235

8.7. Frequency of exposure .....	240
8.8. Usefulness .....	243
8.9. Interviews.....	253
8.10. Summary of findings .....	258
<b>Chapter 9. General discussion .....</b>	<b>261</b>
9.1. Introduction .....	261
9.2. Summary of findings .....	262
9.3. Bringing the evidence together .....	264
9.4. Conclusions .....	267
9.5. Pedagogical implications .....	270
9.6. Limitations and future research.....	273
9.7. Final remarks .....	277
<b>References .....</b>	<b>279</b>
<b>Appendices .....</b>	<b>299</b>



## List of tables

Table 2.1. Mean difference in comprehension percentages between captioning and non-captioning conditions across viewing comprehension studies .....	31
Table 2.2. Studies comparing the effects of subtitles and captions on viewing comprehension from audio-visual input.....	33
Table 2.3. Comparison of the percentage in comprehension with subtitles and captions across studies .....	38
Table 2.4. Studies comparing the effects of subtitles and captions on vocabulary learning through viewing audio-visual input .....	48
Table 4.1. Accumulated lexical coverage per episode .....	96
Table 4.2. Distribution of target items by word families in the BNC/COCA word list..	100
Table 4.3. Frequency of occurrence of TIs per episode and term.....	102
Table 4.4. Distribution of pre-viewing tasks by type and support language .....	110
Table 4.5. Comprehension items by format and type of question .....	116
<i>Table 4.6. Structure of a viewing session.....</i>	<i>126</i>
Table 4.7. Schedule of the intervention.....	129
Table 5.1. Initial proficiency, vocabulary size and familiarity with OV per language condition .....	136
Table 5.2. Distribution of participants for the categorical variables .....	137
Table 5.3. Distribution of items by difficulty index .....	138
Table 5.4. Distribution of items by discrimination index .....	139
Table 5.5. Mean comprehension scores by language condition .....	139

Table 5.6. Independent factors.....	140
Table 5.7. Number of correct and incorrect responses per language condition .....	141
Table 5.8. Results from the GLMM .....	142
Table 5.9. Results from the GLMM: influence of fixed main effects .....	142
Table 5.10. Results from the GLMM: interactions of categorical factors .....	146
Table 5.11. Comprehension scores per episode and language condition .....	150
Table 5.12. Comprehension scores by episode and language condition .....	153
Table 5.13. Top five higher-achievers in the captions condition .....	154
Table 6.1. General proficiency and Vocabulary size descriptive data .....	169
Table 6.2. Total scores in the pre- and post-tests for form and meaning .....	170
Table 6.3. Percentage of relative gains for form and meaning out of 120 target items .....	171
Table 6.4. GLM results for word-form learning .....	173
Table 6.5. GLM results for word-form learning: pairwise contrasts .....	173
Table 6.6. Relative gains in form and meaning per type of instruction and proficiency group .....	174
Table 6.7. GLM results for word-meaning learning .....	176
Table 6.8. GLM results for word-meaning learning: pairwise contrasts .....	176
Table 6.9. Percentage of gains in word-form and word-meaning per term and experimental condition .....	179
Table 6.10. Summary of results from GLMs for form and meaning, by term .....	180
Table 6.11. Correlations between word learning and comprehension .....	182
Table 6.12. Results from regression analysis .....	184

Table 6.13. Scoring in delayed post-test .....	186
Table 6.14. Percentage of relative gains in form and meaning in T3 .....	187
Table 6.15. Percentage of retained word forms and meanings.....	190
Table 7.1. Frequency of occurrence of TIs per episode and term.....	206
Table 7.2. Frequency of encounters with TIs and spacing .....	207
Table 7.3. GLMM results for word-form learning .....	209
Table 7.4. GLMM for word-form learning: pairwise contrasts.....	209
Table 7.5. GLMM results for word-meaning learning .....	210
Table 7.6. GLMM results: effect of categorical variables .....	211
Table 7.7. Word forms learnt by the highest percentage of participants .....	212
Table 7.8. Word meaning learnt by the highest percentage of participants .....	213
Table 7.9. Target nouns descriptives .....	219
Table 7.10. Relative gains for word-form and word-meaning learning in T1 .....	220
Table 7.11. Number (and percentage) of correct and incorrect responses to TNs ....	221
Table 7.12. GLMM results for word-form learning, with CoO .....	222
Table 7.13. GLMM results for word-form learning, with ITOS .....	222
Table 7.14. GLMM results for word-meaning learning, with CoO .....	223
Table 7.15. GLMM results for word-meaning learning, with ITOS .....	224
Table 8.1. Reasons behind the use of on-screen text, by language condition in the intervention.....	239
Table 8.2. Frequency of watching OV input .....	241
Table 8.3. Perceived usefulness for language learning, divided by language condition in the intervention .....	247



Table 8.4. Comparison of perceived usefulness for language learning before and after the intervention.....	249
Table 8.5. Feeling of learning from the intervention .....	249
Table 8.6. Overall assessment of the interventions, general and by experimental condition .....	251
Table 9.1. Cross-sectional comparison of results from Study 1, 2 and 4 .....	264

## List of figures

Figure 4.1. Overview of the intervention.....	86
Figure 4.2. Experimental groups and original sample numbers .....	88
Figure 4.3. Total number of participants in the four studies .....	90
Figure 4.4. Still image of the five main characters in Fresh off the Boat .....	92
Figure 4.5. Checking encounters with TIs across language conditions .....	99
Figure 4.6. Sketch Engine output sample .....	102
Figure 4.7. X_Lex vocabulary size test .....	105
Figure 4.8. Vocabulary test format .....	107
Figure 4.9. Vocabulary delayed post-test (written format).....	108
Figure 4.10. Matching task sample .....	110
Figure 4.11. Word-search task sample .....	111
Figure 4.12. Fill-in-the-blanks task sample.....	111
Figure 4.13. Crossword task sample .....	112
Figure 4.14. Post-viewing task format .....	113
Figure 4.15. Criteria for elaborating the distractors in the post-viewing task.....	113
Figure 4.16. Examples of a MC (top) and TF (bottom) format .....	114
Figure 4.17. Examples of TE and IN formats .....	115
Figure 4.18. Item #1 in BQ1 .....	118
Figure 4.19. Item #2 in BQ1 .....	119

Figure 4.20. Item #3 in BQ1 .....	119
Figure 4.21. Items #4 and #5 in BQ3 .....	121
Figure 4.22. Items #6 and #7 in BQ3 .....	122
Figure 4.23. Items #8 and #9 in BQ3 .....	123
Figure 4.24. Item #10 in BQ3 .....	123
Figure 4.25. Items #2 and #3 in BQ4 .....	124
Figure 5.1. Materials used in Study 1 .....	135
Figure 5.2. Mean comprehension by language and instruction condition .....	143
Figure 5.3. Mean comprehension by vocabulary size .....	144
Figure 5.4. Episodes' mean comprehension by episodes' lexical coverage .....	147
Figure 5.5. Mean comprehension throughout the 24 episodes .....	148
Figure 5.6. Episodes' mean comprehension by session .....	149
Figure 5.7. Episodes' lexical coverage and session.....	152
Figure 6.1. Materials used in Study 2 (part one) .....	167
Figure 6.2. Estimated marginal means for word-form learning: proficiency and instruction .....	175
Figure 6.3. Estimated marginal means for word-meaning learning: proficiency and instruction .....	178
Figure 6.4. Average percentage of gains in form and meaning per session .....	181
Figure 6.5. Word-form learning and comprehension .....	183
Figure 6.6. Word-meaning learning and comprehension .....	183
Figure 6.7. Materials used in Study 2 (part two).....	186
Figure 6.8. Results for word forms in the pre-, post- and delayed-post-test .....	188

Figure 6.9 .Results for word meaning in the pre-, post- and delayed-post-test.....	188
Figure 6.10. Delayed post-test results: word form.....	192
Figure 6.11. Delayed post-test results: word meaning (aural test).....	192
Figure 6.12.Delayed post-test results: word meaning (written test) .....	192
Figure 7.1. Materials used in Study 3 (part one) .....	205
Figure 7.2. Materials used in Study 3 (part two).....	214
Figure 7.3. Comparison of CoO and ITOS measures .....	216
Figure 7.4. Sample of the ELAN software output used to calculate ITOS.....	217
Figure 8.1. Materials used in Study 4.....	233
Figure 8.2. Assessment of variables across questionnaires .....	233
Figure 8.3. Preference for or against on-screen text.....	236
Figure 8.4. Preference for the language of the on-screen text.....	238
Figure 8.5. Frequency of watching OV input before the intervention .....	242
Figure 8.6.Frequency of watching OV input after the intervention .....	242
Figure 8.7. Perception of on-screen text usefulness before and after the intervention .....	244
Figure 8.8. Perceived usefulness for language learning by experimental group.....	248
Figure 8.9. Feeling of learning from the intervention by experimental group .....	251



# List of appendices

## Appendix A

Appendix A. 1. Lexical coverage per episode.....	300
Appendix A. 2. List of target items .....	301
Appendix A. 3. List of non-target items in the vocabulary post-viewing task .....	306
Appendix A. 4. Vocabulary test sample .....	308
Appendix A. 5. Vocabulary pre-viewing task samples .....	310
Appendix A. 6. Vocabulary post-viewing task sample .....	314
Appendix A. 7. Comprehension test sample.....	315
Appendix A. 8. Questionnaire BQ1 (pre-intervention) .....	316
Appendix A. 9. Questionnaire BQ2 (pre-intervention) .....	318
Appendix A. 10. Questionnaire BQ3 (post-intervention).....	321
Appendix A. 11. Questionnaire BQ4 (delayed).....	324

## Appendix B

*(the following documents can be found in the attached CD)*

- Intervention vocabulary tests
- Pre-viewing tasks
- Post-viewing tasks
- Comprehension tests
- Questionnaires



# Chapter 1. Introduction

---

Learning English as a second language (L2) can be particularly challenging in contexts with limited exposure to the foreign language (such as the case of Spain), where opportunities for learning are generally restricted to formal instructional settings. In such contexts, viewing L2 television may be an effective way to increase exposure to authentic spoken language. Television is already an integral part of people's daily life (Webb, 2015), and if language learners were to watch it in the L2 it would be a valuable source of meaning-focused input.

The value of television for L2 learning resides in its potential to provide large amounts of exposure to naturalistic, spoken L2 input, and the fact that language learners are motivated to make use of it as a tool for learning (e.g., Gieve & Clark, 2005). Television presents language in context, and the presence of imagery – which provides an additional semantic support – has been shown to aid information processing. In this sense, L2 television can help learners by increasing the exposure to comprehensible L2 input, which has been shown to be an essential factor in improving second language competence (Ellis, 2013), an even more so for developing listening skills (Vandergrift, 2007) and learners' vocabulary size (Webb & Rodgers, 2009b). Increasing learners' vocabulary can be especially challenging in the EFL context, as there is not enough time in the classroom to teach explicitly the amount of vocabulary needed to understand oral and written discourse (Malone, 2018; Webb, 2015). Watching L2 television may be, therefore, an effective additional resource to obtain this much-needed L2 exposure and develop word knowledge.



The rapid emergence of online streaming platforms in the past years has now endorsed a quick and easy access to a seemingly endless supply of TV programmes and movies, and has made them now available in their original version too – which is often in English. This has provided access to a massive amount of L2 input for EFL (English as a Foreign Language) learners, which is especially useful in a context such as Spain, where there is little presence of English in everyday life outside the classroom. However, while access to L2 television has been increasingly available, Spain remains a traditionally dubbing country, and therefore television is generally consumed in Spanish – or in Catalan, in the case of Catalonia. A shift in viewing habits towards the use of the original version in English would significantly increase the amount of L2 exposure EFL learners receive. However, because television programmes are authentic, unadapted L2 input, EFL learners might be overwhelmed by the fast speech rate, unfamiliar accents and the amount of unknown vocabulary, which would discourage viewing.

As a means to improve the amount of meaning-focused L2 input learners receive, Webb and Rodgers (2009a) proposed the concept of extensive viewing, which involves the regular viewing of TV programmes in and outside the classroom (Webb, 2015). This approach suggests to start L2 viewing within the EFL classroom – providing supported comprehension and adequate-level input – with the aim to encourage learners to engage in autonomous viewing outside the formal setting. Webb (2015) points out several ways to support viewing, such as the addition of captions [L2] or subtitles [L1], the inclusion of pre-viewing activities, and the consecutive viewing of episodes from the same TV series or genre – as they present repeated encounter with unknown words and allow the accumulation of background information, which can aid information processing. Studies on extensive viewing have shown that an extended exposure to comprehensible input contributes to developing listening skills and have potential for incidental vocabulary learning, but research is still needed in this area.

The general aim of this doctoral dissertation is to explore the benefits of an extended exposure to L2 TV programmes with beginner, adolescent EFL learners, with a specific interest in the areas of vocabulary acquisition and viewing comprehension. The study involves a longitudinal classroom-based intervention including 24 consecutive episodes of a TV series, and it examines the potential benefits of different types of viewing support. It is divided into four main studies, each one focused on a different aspect of language learning, examining the potential benefits of both subtitles in the native language and in the target language, and the differential effects of having explicit instruction on vocabulary or not. It also aims at investigating the role of repetition and imagery on word learning, and evaluating the long-term effects that this type of input may have on learners' attitudes and viewing habits.

This dissertation starts with a comprehensive literature review of the theoretical background in this area of research (Chapter 2), followed by a brief rationale and the presentation of the main research questions that each one of the four studies will address (Chapter 3). Then, a methodology section (Chapter 4) provides an overview of the study design, including a detailed description of the audio-visual materials (i.e., the TV series selected) and the testing instruments, and the procedure followed during the 10-month long intervention. The next four chapters present the results from the four studies conducted. For clarity reasons, each of the studies starts with a brief summary of the relevant gaps in that particular research area and a short review of their specific instruments and methodology. Study 1 (Chapter 5) examines the differential effects of captions and subtitles on learners' viewing comprehension of the series, and looks into several mediating factors that may also play a role in comprehension, such as the episodes' lexical coverage. Study 2 (Chapter 6) examines vocabulary learning through the TV series under two instructional conditions, and investigates the relationship between learner's proficiency and vocabulary gains. Long-term retention of the words

learnt during the intervention are examined as well, and the relationship between vocabulary gains and comprehension rates (in Study 1) are also explored. Study 3 (Chapter 7) investigates word-level characteristics, including frequency and spacing, with a special interest in the effect of the imagery associated with videos as a support for vocabulary learning. Study 4 (Chapter 8) dives into learners' attitudes towards using audio-visual media for language learning purposes, reports on learners' feeling of learning from the classroom intervention, attention to and enjoyment from the series, and explores whether learners' perceptions have changed over time. Finally, the last chapter (Chapter 9) brings together the results from the four studies involved in this dissertation, presents a summary of the main findings, acknowledges the limitations of the studies, discusses their pedagogical implications and suggests directions for further research in the area.

## **Chapter 2. Literature review**

---

This chapter addresses the multiple theoretical and empirical sources of this dissertation. After a brief introduction to multimedia learning, it delves into audio-visual input as a specific case of multimedia input. Then, it addresses the features that makes this type of input suitable for second language learning, and the differential benefits brought by the addition of captions and subtitles. Next, it focuses specifically in comparative studies on comprehension and vocabulary learning, and discusses several mediating variables that have been shown to play a role in language learning through audio-visual input.

### **2.1. Introduction**

Nowadays it is commonly accepted that audio-visual materials (with or without on-screen text support) are robust tools for language learning because they expose the learner to a large amount of rich, authentic input (Danan, 2004). Unlike artificial material designed specifically for language learning purposes, television, movies and other authentic video materials (e.g., YouTube videos) provide language learners with a source of naturalistic oral language input that resembles real life, as the images and contextual clues make it possible to “view” the message as well as listen to it (Baltova, 1994; Danan, 2004). The presence of visual information provides an additional semantic information that aids comprehension and potentially supports vocabulary learning (e.g., Rodgers, 2016, 2018a). The simultaneous presentation of visual and verbal stimuli activates two different information-processing channels that interact, resulting in a greater depth of spoken-word processing (Bird & Williams, 2002). This, in turn, would

explain why language learning can be enhanced through multimedia learning – by combining images with verbal information (Montero-Perez, Noortgate & Desmet, 2013; Plass & Jones, 2005; Sydorenko, 2010).

Multimedia learning can be succinctly defined as learning from words (spoken or written) and pictures (static or dynamic) (Mayer & Moreno, 2003). The main theoretical foundation in multimedia learning is Mayer's (2001) Cognitive Theory of Multimedia Learning, which integrates previous cognitive theories of learning such as Paivio's (1986) Dual Coding Theory and Chandler and Sweller's (1991) Cognitive load Theory. The Cognitive Theory of Multimedia Learning is based on three main assumptions on how the human mind works: (1) humans process information through two different systems that interact: the verbal/auditory channel and the visual/pictorial channel (the dual-channel assumption); (2) humans' capacity to process information simultaneously through the verbal and visual channels is limited (the limited capacity assumption); and (3) meaningful learning requires a considerable amount of cognitive processing to integrate verbal and visual information into the existing knowledge (the active-processing assumption) (Mayer & Moreno, 2003).

Research in multimedia learning has provided several evidence-based principles to take into account in the design of multimedia instruction<sup>1</sup>, such as the split-attention principle (i.e. people learn better when words and pictures are physically and temporally integrated) or the redundancy principle (i.e. people learn better when information is not presented concurrently in multiple forms) (Mayer, 2014). Multimedia learning theories are, however, a universal approach to learning, and are not specific to second language learning (SLA) context. While there has been research in the area of multimedia for

---

<sup>1</sup> Multimedia instruction refers to using the combination of words and pictures with the intention to promote learning.

language learning (e.g., Plass & Jones, 2005), findings are still unclear, for instance, regarding the use of video enhanced with on-screen text (Sydorenko, 2010). These theories provide, however, a theoretical foundation for the use of multiple input modalities in SLA.

Early research on multimedia input found that, compared to traditional audio input (e.g., listening to an audio-recorded conversation), audio-visual input enhances foreign language processing (Meskill, 1996), and helps develop listening skills (Brett, 1996) and vocabulary acquisition (Brett, 1998). Research on the use of multimedia for language learning purposes has gone well beyond asking whether to use or not use the video, but rather how to present the audio-visual materials in an efficient way (Baltova, 1994; Montero-Perez et al., 2013).

## **2.2. Audio-visual input and second language learning**

Audio-visual input can substantially support second language (L2) learning by providing learners with comprehensible input. Indeed, increasing the amount of exposure to comprehensible input in the target language is beneficial for second language acquisition (Ellis, 2013), even more so for developing listening competence – an often overlooked skill in the language classroom (Nation & Newton, 2009; Vandergrift, 2007). Improving the understanding of oral discourse is one of the most difficult challenges L2-learners encounter (Graham, 2006; Hasan, 2000; Kim, 2015) especially when they find themselves in an environment – such as the case of Spain – where they are not regularly exposed to the target language. In this context, it seems especially relevant to investigate ways to expose language learners with suitable, meaningful input, as “listening comprehension is at the heart of L2 learning; and the development of L2 listening skills

has demonstrated a beneficial impact on the development of other skills (e.g., Dunkel, 1991; Rost, 2002)” (Vandergrift, 2007: 191).

A case in point is the attainment of a large enough vocabulary size in the L2, which has been shown to be a strong predictor of comprehension (e.g., Montero-Perez, Peters, Clarebout & Desmet, 2014; Rodgers, 2013). Research has shown that learners need to know around 3,000 word families to understand oral discourse (e.g., van Zeeland & Schmitt, 2013) and between 8,000 and 9,000 to understand written discourse (e.g., Nation, 2006). However, in classroom settings, the amount of time that can be devoted to vocabulary learning is limited, and there is a sizable gap between the amount of words that can be explicitly taught and learnt in class and those necessary to achieve higher second language (L2) proficiency (Malone, 2018). Research has shown that extensive reading outside the classroom can foster vocabulary acquisition (Nation, 2015; Schmitt, 2008), but its impact remains limited: the typical learner does not read as much as to encounter the same words frequently enough to avoid forgetting them (Laufer, 2005). Additionally, there has been a drop in popularity in reading habits (European Commission, 2017), especially among young people, who prefer watching TV to reading (Lindgren & Muñoz, 2013; Peters, 2018), a situation also found in countries such as Canada and the United States, where people watch TV five times more than they read (Statistics Canada, 2005; United States Department of Labor, 2006). While it might seem quicker to read a book, people spend more time watching TV than reading, and thus, could potentially learn more from this type of input (Rodgers, 2013).

In view of this, Webb and Rodgers (2009a) proposed extensive viewing as an alternative way to increase the amount of L2 exposure, as it provides a source of rich authentic input in an environment that has limited L2 presence (Webb, 2015). TV programmes have been shown to be an effective source of comprehensible input and of natural, contextualized spoken dialogue (Vulchanova, Aurstad, Kvitnes & Eshuis, 2015),

with the additional semantic support of the images (Rodgers, 2013). This type of input has also showed potential to learn vocabulary incidentally due to its lexical richness and repeated encounters with low frequency words (Rodgers, 2018a; Sydorenko, 2010; Webb & Rodgers, 2009b). The emergence of multimedia learning environments and the greater accessibility to series, movies, and other online streaming platforms in recent years have created opportunities for teachers and learners to boost language learning inside and outside formal settings.

Watching television has been found to be one of the most common self-directed learning strategies amongst European students, who are also more likely to learn independently through this media than doing so through extensive reading (Gieve & Clark, 2005). The benefits brought by audio-visual materials, however, can also be additionally boosted within the FL classroom by adding pre-viewing activities to direct learners' attention to lexical items, which can aid word learning and ease comprehension. Most research in the area has been focused on the advantages of audio-visual input for incidental vocabulary learning, with adult learners and outside the classroom context. Fewer studies have looked at younger, less proficient learners, who might also need additional support when confronted with authentic input. More research in this area is needed, especially regarding the combination of this type of input with explicit instruction and younger population.

### **2.2.1. Suitable input for L2 learners**

TV programmes – along with other audio-visual materials such as films, documentaries, or short videos – comply with Nation's (2007) five conditions for suitable input (Rodgers, 2013): it is processed in large quantities, it is comprehensible, it is engaging (Webb, 2010a), it is familiar to the language learners, and it provides contextual



cues (i.e. through image and dialogue) (Rodgers & Webb, 2011). TV programmes also have several features that can contribute to the facilitation of information processing, the main ones being the presence of imagery, the possibility to accumulate background knowledge and the recurrence of vocabulary, as will be discussed in the following sections.

**Input has to be processed in large quantities.** TV provides a large amount of input on a short time, and it is already consumed in large quantities across the EU, with 81% of the population watching it daily (European Commission, 2017). This figure goes up to 88% in Spain, a traditionally dubbing country, which means that most of this input is in Spanish (Almeida & Costa, 2014) rather than in the OV (Original Version), generally in English. Compared to other European countries – in which learners are frequently exposed to English through television, movies or newspapers (e.g., Vulchanova, et al., 2015) – in Spain most of the exposure to English is limited to formal instructional settings (Muñoz, forthcoming), where there is not enough time to provide learners with as much exposure to the L2 as needed. If L2 learners were to watch TV in the L2 for enjoyment – even if just a portion of it –, it could be a valuable source of meaning-focused input (Webb & Nation, 2017). The appearance and rise of online TV streaming platforms such as Netflix or HBO grants viewers quick and easy access to a wide variety of TV series in the English original version (OV). These types of platforms are changing the way people consume TV and facilitate binge watching a large amount of movies and TV programmes of their choice. In the case of TV series, episodes are rarely viewed in isolation, and when someone enjoys a series, they would normally choose to watch not one episode but the entire season(s), which would ensure even more exposure.

**Input has to be comprehensible.** We acquire language by understanding the messages that we receive (Krashen, 1985), and materials should be selected in a way that matches learners' proficiency level (Webb, 2015). Studies on reading and listening

have provided estimates of the amount of vocabulary needed for adequate comprehension – as prior vocabulary knowledge has been shown to be a strong predictor in comprehension –, and suggested that learners need a lexical coverage (see below) ranging from 90% to 99% (Hu & Nation, 2000; Laufer & Ravenhorst-Kalovski, 2010; Noreillie, Kestemont, Heylen, Desmet & Peters, 2018; Stæhr, 2009; Van Zeeland & Schmitt, 2013; Webb & Rodgers, 2009b). Webb and Rodgers (2009b) analysed the scripts of over 300 movies and found that the knowledge of the most frequent 3,000 word families already provided 95% coverage. Although knowing this amount of vocabulary does not guarantee adequate comprehension, Webb and Rodgers suggest that comprehension could be improved as learners kept on watching more movies. The presence of imagery would also aid comprehension, and the access to comprehensible input could be improved by the addition of captions or subtitles, as a means to facilitate viewing comprehension.

**Input has to be engaging.** Because of its popularity as a leisure activity, it seems safe to assume that watching TV series is indeed engaging. The wide variety of TV genres and diverse range of topics available nowadays makes it easier to simply drop a series when it is not of your interest, and quickly pick up a new one. Furthermore, studies that have investigated attitudes towards watching videos for language learning have generally found that learners considered it a positive and engaging activity (e.g., Chung, 1999; Rodgers, 2013; Steward & Pertusa, 2004), and that they feel that videos were easier to understand compared to other types of input (Chung & Huang, 1998; Gruba, 2006, Hasan, 2000). Additionally, language learners are highly motivated to watch visual media for language learning (Webb, 2010b), and watching TV series has been reported as a common autonomous learning strategy amongst FL learners (Chapple & Curtis, 2000; Gieve & Clark, 2005).

**Input has to be familiar.** Learners have to be familiar with the content of the input to be able to attain an adequate level of comprehension. One feature of TV series in particular that facilitates comprehension is that TV series are sequential in nature – compared to other audio-visual materials, such as educational videos –, and (as mentioned above) people tend to watch more than a single episode (Rodgers, 2013). Learners can increase their familiarity with the setting, the characters and even the language of a specific TV series by viewing multiple and successive episodes, as well as getting used to the accents. Watching series from the same genre would also facilitate being acquainted with specific vocabulary, while watching different genres would provide the opportunity of accessing new lexical fields.

**Input has to be contextualized.** Learners should be able to gain knowledge of vocabulary using contextual cues and background knowledge (Nation, 2007). The imagery and dialogue present in TV provides a contextualized setting in which learners encounter new words and expressions in an authentic context and see them used in real-life situations. The lack of contextual knowledge can be an obstacle for comprehension, and in real life interactions, people normally have some sort of knowledge of a situation and interlocutors before engaging in a conversation (Hasan, 2000). The episodic nature of TV has the potential of accumulating and increasing language knowledge through viewing multiple episodes of the same series (Webb, 2011). Research on incidental vocabulary learning has also suggested that the more the learner encounters unknown words in context, the more changes they have to learn those words (e.g., Horst, Cobb & Meara, 1998; Waring & Takaki, 2003; Webb, 2007; Webb & Rodgers, 2009a).

### **2.2.2. Narrow viewing**

One of the conditions for suitable input is that input has to be familiar. In real-life situations, we usually would have some background information about our interlocutor and the topic that is going to be discussed. The lack of contextual knowledge and vocabulary in a decontextualized situation (such as a traditional listening task) can hinder the ability to predict and, therefore, obstruct comprehension (Hasan, 2000). TV programmes are serial in nature and episodes are rarely watched in isolation, which allows the viewer to gather background information as he or she keeps on watching. On the basis of the findings from narrow reading (e.g., Schmitt & Carter, 2000), which show that reading texts with similar topics and plot lines may facilitate comprehension, Rodgers and Webb (2011) proposed the concept of narrow viewing.

Narrow viewing involves watching successive episodes of the same TV programmes in chronological order, which would allow learners to gradually accumulate background knowledge and facilitate comprehension of the content (Webb, 2015), another factor that may aid top-down processing. The more episodes you watch from the same TV series, the more information you have about the reoccurring characters and their dynamics, making it possible for the viewer to predict or guess what a character will do or say as they accumulate information. Successive viewing of the episodes would also help viewers to get used to accents or the way characters talk (e.g., Rodgers, 2013; Vanderplank, 2019). Additionally, it has also been found that television programmes on the same topic or genre have a smaller vocabulary load (Rodgers & Webb, 2011), which would suggest that targeting programmes of a similar nature (as opposed to watching single episodes randomly) would have more potential for learning. The advantages of having prior information for comprehension have also been demonstrated by research

on listening comprehension and advanced organizers<sup>2</sup> (e.g., Chung, 1996, 1999), as will be discussed below.

Webb (2010a), and Rodgers and Webb (2011) investigated the re-occurrence of vocabulary in movies and related TV programmes. Webb (2010a) analysed the scripts of 143 movies and found that words did not re-occur often within the same movie, and only a few were encountered more than 10 times – which suggested that if learners watched only few movies, the number of words learnt would be small. However, as the number of movies analysed increased, the number of words encountered more than 10 times also increased. Rodgers and Webb (2011) also found that related television programmes are likely to contain fewer word families than unrelated programmes, and that word families from the 4,000–14,000 levels were more likely to reoccur in a complete season of a television programme than in a random sample of programmes. The repeated encounters with low frequency words within one specific TV programme suggest that the more episodes you watch from the same TV series, the more potential to learn from them (Webb, 2011), which is what people would normally do – watching complete seasons of a TV series rather than isolated, non-consecutive episodes (Rodgers, 2013). However, most studies investigating audio-visual input for L2 learning have used short clips, segments of films or educational videos, which are largely unrelated, while only a few studies have used several, full-length TV episodes to investigate comprehension and vocabulary learning (e.g., Bravo, 2008; Chen, Li & Todd, 2018; Frumuselu, Maeyer, Donche & Gutiérrez, 2015; Gesa, 2019; Rodgers, 2013; Rodgers & Webb, 2017; Zarei, 2008).

---

<sup>2</sup> Advanced organizers can be generally described as activities given prior to a listening or viewing activity as an introduction.

### 2.2.3. Imagery as an aid for comprehension and L2 learning

Compared to traditional listening activities, the first and most obvious attribute of audio-visual materials is imagery, which, as a semiotic resource, is a powerful mode of meaning-making (see The Douglas Fir Group, 2016). Listening comprehension is an “active cognitive process” involving “speculating and predicting” rather than deciphering individual sounds as listeners often need to fill in missing information in an incomplete acoustic input (Noblitt, 1995). Vandergrift (2007) gives a succinct explanation of the listening process:

“Understanding spoken language is essentially an inferential process (Rost, 2002); linguistic knowledge and world knowledge interact in parallel fashion as listeners create a mental representation of what they have heard (Hulstijn, 2003). Listeners apply these knowledge sources using top-down and bottom-up processes (Lynch & Mendelsohn, 2002; Rost, 2002; Flowerdew & Miller, 2005). Listeners favour top-down processes when they use context and prior knowledge (topic, genre, culture and other schema knowledge stored in long-term memory) to build a conceptual framework for comprehension. Top-down processes are developed through practice in the use of compensatory strategies. Listeners favour bottom-up processes when they construct meaning by accretion, gradually combining increasingly larger units of meaning from the phoneme-level up to discourse-level features. Bottom-up processes are developed through practice in word segmentation skills.” (Vandergrift, 2007: 193)

Comprehension occurs when sufficient information is acquired through both types of processing (Buck, 2001). In the context of television viewing, the presence of imagery allows viewers to construct meaning using an additional source of non-verbal information (the images). Research has shown that low-proficiency listeners rely more on top-down processing, because of their limited word segmentation skills and proficiency. Imagery – which can work as a “compensatory mechanism” (Vandergrift, 2007: 193) – provides

beginner-level learners with contextual knowledge which allows them to shift attention away from top-down processing and focus their attention on details of the story, which in turn can positively affect comprehension (Rodgers, 2013, 2016).

The benefit brought by the addition of visual input to aural input for comprehension was explained by Paivio's (1986) Dual Coding Theory, which argues that verbal stimuli (language) and non-verbal stimuli (image) are processed by two different systems that interact. The activation of both systems results in better recall and greater depth of processing – which would explain why L2 learning can be enhanced by combining images with verbal information (Sydorenko, 2010). Imagery has also been shown to have an advantage over verbal input for information processing (Paivio, 2008), and it leads to deep processing of aural input, especially when combined with text (e.g., Jones, 2003). This can also be linked to Mayer's (2001) Cognitive Theory of Multimedia Learning, according to which people learn more deeply through the combination of words and images.

A number of studies support the idea that language learners can use the imagery associated with videos to assist information processing, and findings from research on listening and viewing have shown that the presence of images has a positive impact on comprehension, as images allow the viewer to see the message as much as to listen to it. Jones and Plass (2002) explored comprehension of listening to a historical text with and without access to images, and found that learners with access to them scored significantly higher than those without access to the visual aid. A similar result was found by Maleki and Safaee Rad (2011), who compared performance in items in IELTS test with and without the support of images. They found, again, that access to the images led to higher comprehension, but that differences were only significant when proficiency was low.

The importance of images for comprehension seems especially important, therefore, during the first stages of comprehension. Gruba (2004, 2006) investigated how learners make use of the images available in video fragments through retrospective verbal reports. While he found that learners make use of both the aural and visual input for comprehension, Gruba reported that visual elements influence on comprehension shifts from primary to secondary importance as the listeners' develop understanding of the text. Thus, "listeners first developed macrostructures based on visual elements and these guided them in their initial understandings of the videotext. As their comprehension matured, aural elements gained prominence as they sought detail and clarification" (Gruba, 2004: 76). In a study about learners listening strategies, Hasan (2000) found that learners reported difficulties in understanding listening texts because they could not see the speakers' body language, faces and gestures. Visual clues – in form of pictures, video or writing – helped them understand the content, while it also helped reduce anxiety when confronting unfamiliar topics. In a study by Ockey (2007) it was observed that learners used more visual clues from videos than from still images, though both were a helpful support while listening. Baltova (1994) also reported the positive effects of visual clues from videos, and found that learners who had access to audio and video almost doubled the comprehension scores of the audio-only group. Results seem to indicate that comprehension increases when viewers have access to images related to the information presented in the audio, and that this is especially true for learners with a low proficiency level. Sydorenko (2010) examined participants' insights from viewing captioned clips and found that, while participants paid more attention to the captions than to the video, most of them reported that the video was more helpful than captions, especially for beginner learners. The author suggested that while learners could understand only part of the L2 input, they had no difficulties in processing the image.



While it has long been hypothesised that the images in the video might have a positive effect on vocabulary learning, research on vocabulary learning through audio-visual input has just recently started to investigate to what extent image supports aural information. Rodgers (2018a) compared the degree to which the aural form of 90 target words co-occurred simultaneously with the presentation of its image in single seasons of television, comparing two types of genres (narrative TV and documentary). Following on the temporal contiguity principal of multimedia learning – which states that students learn better when words and pictures are presented simultaneously rather than successively (Mayer, 2014) –, it can be assumed that, for a learner to be able to use the images to infer the meaning of the unknown word, the image associated with that word should appear in close proximity with its aural form. As Rodgers argues, “the temporal proximity allows a learner to make fewer demands on their mental processes enabling them to hold the separate representations (aural and visual) in their working memory to better build mental connections between them (Mayer, 2001)” (Rodgers, 2018a: 197). Results showed that the imagery found in documentaries potentially supported vocabulary learning more than narrative TV, with 65% of the images co-occurring simultaneously with the target aural word forms, and over 70% co-occurring within a 10-second timeframe. In contrast, in the narrative programme, only 29% of the target items were found to have image support. Findings indicated that the way in which the visual and aural form of words co-occurs may support vocabulary learning through viewing television, although more research is needed to investigate the extent to which co-occurrence affects vocabulary learning.

In a pilot study including participants from the current sample, Pujadas and Muñoz (2018) investigated the effect of imagery co-occurrence and image time on-screen on vocabulary learning through extensive TV viewing with adult and young learners at different levels of proficiency. They found that for older and more proficient learners,

words that occurred simultaneously in the image and the audio were better recalled. For younger and less proficient learners, however, words whose image appeared on screen for a longer time were better recalled – independently of co-occurrence with the aural form –, suggesting that “imagery can function as a ‘sketch pad’ for holding information to be worked on by other cognitive mechanisms” (Paivio, 2008: 28). More recently, Peters (2019) investigated the effects of image support on incidental vocabulary learning through a documentary, and found that words that had a visual representation on-screen close to the aural occurrence were more likely to be learnt. These studies suggest, firstly, that images provide some sort of on-line visual semantic support – which appears to vary across genres – and that that support has a direct impact on vocabulary learning. Secondly, individual differences may affect how learners make use of this additional information.

### **2.3. Captions and subtitles**

In spite of the potential benefits of audio-visual input for language learning purposes, TV series might be too demanding for those learners whose L2 language skills are not high enough to achieve satisfactory comprehension (Webb, 2010b; Webb & Rodgers, 2009a). Because of their limited linguistic knowledge, beginner-level learners cannot process input automatically as more advanced learners do (Vandergrift, 2007). They have to consciously decode aural input into meaningful units (bottom-up processing), and “a large proportion of what they hear may be lost, given the speed of speech and the inability of working memory to process all the information within the time limitations” (Vandergrift, 2007: 193). A number of factors can affect this process, including learners’ ability to recognize words and recall their meaning (Buck, 2001) and learners’ prior vocabulary knowledge (e.g., Webb & Rodgers, 2009b).

Given the need to provide learners with non-adapted, natural samples (Gilmore, 2007), the addition of on-screen text may provide access to authentic foreign language material that would otherwise be difficult to comprehend for non-native speakers (Vanderplank, 2016a). There are several types of on-screen text, according to the language and format. *Captions* (also referred as intralingual subtitling, bimodal subtitling or same-language subtitling in research literature) are presented in the same language as the audio (e.g., English audio + English text). They were primarily used in the context of the hard-of-hearing population, and are a word-by-word transcription of the soundtrack. *Subtitles* (also referred as interlingual subtitling, standard subtitling or traditional subtitling) typically refers to on-screen text in the native language of the viewer and provide a translation of the foreign audio (e.g., English audio + Spanish text). The terms captions (L2 audio + L2 text) and subtitles (L2 audio + L1 text) will be used throughout this literature review and the subsequent presentation of the studies. Other types of on-screen text that can be found in language learning studies include reversed subtitling (L1 audio + L2 text), key-word captioning (L2 audio + L2 key-words), enhanced captioning (L2 audio + L2 full-captioning with highlighted words) and dual subtitling (L2 audio + L1 and L2 text simultaneously). The following sections will be mainly focused on full-captioning studies (either captions or subtitles), since this is the format generally available for L2 learners.

### **2.3.1. Research on captioned and subtitled audio-visual input**

Before diving into the benefits of captions and subtitles, it is worth to consider the effects that adding on-screen text has on learners' attention and processing of the input. In previous sections it has been discussed how imagery aids processing and provides semantic support. The addition of text also alters the way viewers process information,

and can be a valuable addition for language learning. However, language teachers and students are sometimes reluctant to use captioned or subtitled material.

In classroom settings, a concern for many teachers is that learners might not attend to the audio when they also have captions or subtitles, which would hinder their listening skills development (Borras & Lafayette, 1994). Research has found that the reading of the on-screen text is automatic, as demonstrated by eye-tracking studies (e.g., Bisson, Van Heuven, Conklin & Tunney, 2014; d'Ydewalle, Praet, Verfaillie, & van Rensbergen, 1991). Captions and subtitles are consistently read – whether the sound is on or off, and independently of participants' proficiency level or their familiarity with viewing audio-visual materials in a foreign language (d'Ydewalle & Gielen, 1992; Pavakanum & d'Ydewalle, 1992). However, automatic reading of the on-screen text does not prevent viewers from processing the audio (Bird & Williams, 2002; Garza, 1991; Markham, 1999; Price, 1983), nor the image (Bisson et al., 2014). Anecdotal reports of incongruities between text and sound also support the idea that both channels can indeed be processed simultaneously (Danan, 2004; Garza, 1991). Studies gathering learners' insights have also shown, however, that learners tend to pay more attention to the text than to the image or the audio (Sydorenko, 2010).

It has been argued that, when both text and sound are present attention is “divided between the two depending on viewers' needs, with more time usually devoted to subtitles to process complex information” (Pavakanum & d'Ydewalle, 1992), although there is also evidence that beginner learners tend to prioritize the image, which can be processed regardless of the learner' L2 skills (Sydorenko, 2010). Changes in on-screen text reading behaviour can also be observed when images are more salient, as they capture viewers' attention (Perego, Del Missier, Porta & Mosconi, 2010), suggesting that sometimes a competition exist between the two sources of information. In Bisson et al. (2014), eye-tracking data revealed that learners did not use the full time the text was

available on screen to read it, and that they returned to the image instead. This supports results from prior studies indicating that “the reading of subtitles still allows for the processing of the images (Perego, Del Missier, Porta, & Mosconi, 2010) instead of being just a reading exercise (Jensema, El Sharkawy, Danturthi, Burch, & Hsu, 2000).” (Bisson et al., 2014: 15).

The reason why learners might not pay attention to audio is that they have to divide their attention among three types of stimuli – visual, aural and text – and aural stimuli seems to be less salient. As attentional capacity is limited, learners have to use attention selectively (Robinson, 2003), to avoid a cognitive overload. The redundancy principle (Mayer, 2014; Sweller, 2005) assumes that redundant material slows down information processing and learning. However, research suggests that, even if for native speakers content learning may be easier without the on-screen text, it might be claimed that for L2 learning three modes of presentation are more beneficial than two (Sydorenko, 2010). Even if it is cognitively more demanding, processing of multiple modalities does not seem to hinder language learning. The advantage of redundancy through multiple modalities can be attributed to learners’ individual preferences, as it is the learner who chooses what to focus on. The availability of video, audio and text allows learners to decide where to direct their attention to and process information from, according to their own preferences (e.g., visual or verbal information) and their language skills. Having on-screen text support could be more useful than distracting because it helps learners parse the stream of speech “into chunks which reliably mark meaning”, an essential step in “learning to understand a language” (Ellis, 2003: 77).

The bulk of research in the field has been focused to the use of captioning, and especially in the processing of spoken language (Danan, 2004). Since Price’s trailblazing work in 1983, studies on L2 learning from captioned and non-captioned audio-visual materials have consistently shown the overall advantage of viewing videos enhanced

with on-screen text compared to viewing them without it for both content comprehension and vocabulary learning (e.g., Baltova, 1999; BavaHarji, Alvi & Letchumanan, 2014; Chai & Erlam, 2008; Chen, Liu & Todd, 2018; Chung, 1999; Danan, 1992; Garza, 1991; Guillory, 1998; Hsu, Hwang, Chang & Chang, 2013; Huang & Eskey, 1999; Li, 2014; Mohd Jelani & Boers, 2018; Montero-Pérez, Peters & Desmet, 2014, 2018; Montero-Pérez, Peters, Clarebout & Desmet, 2014; Rodgers, 2013; Rodgers & Webb, 2017; Sydorenko, 2010; Winke, Gass & Sydorenko, 2010). The advantages of captioning have also been shown independently of the type of captions accompanying the video (e.g., full captioning, key-word captioning, reversed subtitling<sup>3</sup>) or type of measures used to assess incidental vocabulary gains (e.g., recognition, recall) or comprehension (e.g., multiple-choice test, essay) – although there is considerable variation in the outcomes due to testing differences.

Similar findings are shown by studies including subtitles in the learners' native language, with the subtitles condition also outperforming no-text conditions. However, whether it is captions or subtitles that are more useful in general in the language-learning context, and especially for learners with limited L2 skills, is still a matter of debate, with studies showing mixed results depending on learners' proficiency and the language aspect that is being assessed. While captions have consistently been shown to better aid various aspects of word learning (e.g., Charles & Trenkic, 2015; Markham, 1999; Mitterer & McQueen, 2009; Sydorenko, 2010; Winke et al., 2010), the majority of studies on comprehension concur that subtitles facilitate understanding of the content (e.g.,

---

<sup>3</sup> Several of the aforementioned studies also include reversed subtitling (L1 sound, L2 text) or key-word captioning, but these types of on-screen text support will not be addressed in this section. Although research has shown that reversed subtitling also has advantages for language learning purposes (e.g. Baltova, 1999; Danan, 1992), the presentation of audio-visual material with L1 audio is not a common practise in either classroom setting nor in viewing TV for entertainment. Therefore, considering the scope of the present study, they will not be further discussed.

Bianchi & Ciabattoni, 2008; Birulés-Muntané & Soto-Faraco, 2016; Latifi, Mobalegh & Mohammadi, 2011; Markham, Peter & McCarthy, 2001; Markham & Peter, 2003).

Finding a balance between understanding the content and learning vocabulary seems especially relevant in the context of the EFL classroom, where both objectives are of equal importance. The distinctive benefits of both types of on-screen text will be discussed in the following sections.

### **2.3.2. Advantages of captioning**

Numerous studies have addressed the diverse benefits of captions for second language learning. Captions make content more accessible when the language of the authentic materials lays beyond students' level of proficiency (Guillory, 1998), and there is substantial evidence of the value of captions in training listening skills (Vanderplank, 2014). In addition to comprehension, captions help vocabulary building (Danan, 2004). Research has found that the presence of captions encourages the development of word decoding and segmentation skills, aiding written form recognition by allowing listeners to visualize the speech stream, and facilitating the identification of word boundaries (Baltova, 1994; Bird & Williams, 2002; Markham, 1999; Neuman & Koskinen, 1922; Sydorenko, 2010; Winke et al., 2010). They can also aid in stream chunking problems (Graham, 2006), and reduce listeners decoding efforts (Buck, 2001). Captions also aid aural form recognition (Markham, 1999), as they allow listeners to create a phonological visualization of the aural cues in their mind, making ambiguous words clearer and improving listening-based recognition of those words when they are subsequently heard (Bird & Williams, 2002; Markham, 1999). Learners can then form a memory trace of the words more accurately, and identify more easily identical sounds without text support later (Bird & Williams, 2002; Garza, 1991).

Bimodal presentation also helps connecting auditory to visual input (Garza, 1991) and encourage sound-script automatization (Goh, 2000), which may support learners in constructing an initial form-meaning mapping (Doughty, 2004; Winke et al., 2010). Making this connection in the mental lexicon is an essential part of the language learning process and a crucial step in vocabulary acquisition (Laufer, 2005; Montero-Perez, Peters, Clarebout & Desmet, 2014). For vocabulary learning, captions may also contribute to a “conscious focus on the form (especially the correct form), particularly when new or striking expressions are used” (Vanderplank, 1988: 276), which in turn may encourage deliberate attention, which is crucial to “taking out” language (Vanderplank, 1990). By helping learners to better isolate individual words, they can also reinforce clues for inferring word meaning and aid building vocabulary (Danan, 2004; Montero-Pérez, Noortgate & Desmet, 2013), and help learners produce newly acquired words by themselves (Baltova, 1999; Danan, 1992). The presence of the L2 text also allows listeners to understand details that otherwise would be missed (Danan, 2004), and may aid comprehension by facilitating additional cognitive processes, such as greater depth of spoken-word processing (Bird & Williams, 2002).

Overall, captions may be seen as a mediating device (Vanderplank, 1988), as they help learners visualize what they hear, especially when input is slightly beyond their actual linguistic abilities (Danan, 2004; Markham, 1989; Montero-Pérez, Peters & Desmet, 2014; Neuman & Koskinen, 1992). Captions can also make a huge difference when watching a comedy, which relies on rapid verbal interaction for much of its humour; and native speakers can also find them helpful when there are strong or regional accents (Vanderplank, 2016a).

Captions bridge the sizable gap between reading comprehension and listening comprehension skills, while improving both at the same time: “By providing students with a familiar (i.e. comprehensible) graphic representation of an utterance, they are



empowered to begin to assign meaning to previously unintelligible aural entities, gradually building their aural comprehension in relation to their reading comprehension” (Garza, 1991: 246; also Borrás & Lafayette, 1994; Danan, 2004; Garza, 1991; Markham & Peter, 2003). They enhance learning by (1) allowing learners to transfer their developed reading skills to help strengthen/develop aural comprehension, (2) increasing accessibility of the salient language and giving the opportunity to enjoy same input understood by a native-speaker, (3) allowing use of multiple language processing strategies to process multiple modalities of input, (4) increasing the memorability of essential language, and (5) promoting the use of vocabulary in appropriate context (Garza, 1991).

For beginner learners with better reading than listening skills, captions facilitate recognition of written word forms while audio facilitates recognition of aural word forms (Sydorenko, 2010), and they can be, additionally, a reading practise (Koolstra et al., 2002). However, “when you add captions as a supposed support – which assumes the ability to read and understand at a reasonable speed – you introduce a reading speed variable” (Vanderplank, 2016b: 85). The usefulness of captions seems, therefore, heavily dependent on the learners’ reading ability and aural decoding skills; and arguably the speed of the dialogue is something that is not controlled by the viewer (Baltova, 1999; Danan, 2004). The impact of captioning on learning also depends on the gap between the students’ proficiency level and the difficulty of the spoken text, as captions cannot compensate for an excessive wide gap (Bianchi & Ciabattini, 2008; Guillory, 1998), and input that is too far beyond the linguistic competency of the learners may yield poor language gains (Montero-Perez, Peters & Desmet, 2014).

### **2.3.3. Advantages of subtitling**

As seen above, in spite of their benefits, captions may not be suitable for all learner profiles (Danan, 2004), especially for beginner-level students or younger learners, with limited L2 language skills, as captions do not compensate for fast speech and unknown vocabulary (Guillory, 1988). Neuman and Koskinen (1992) found that students with limited linguistic abilities learned less from captions than students with higher proficiency. Similarly, Lambert and Holobow (1984) reported that captions had limited benefits compared to forms of subtitles that incorporated native language input. This suggests that a minimum L2 competency threshold might be necessary to benefit from captioning (Danan, 2004; Neuman & Koskinen, 1992; Lambert & Holobow, 1984; Winke et al. 2010).

Per contra, subtitles allow understanding of the input regardless of the viewers' proficiency level. Findings support Krashen's "comprehensible input" theory (Krashen, 1985), which claims that we acquire language by understanding the messages that we receive. If the input is not understood, learning is unlikely to occur, since learners may not pay attention to it. This is also supported by Muñoz's (2017) eye-tracking study, which revealed that young, low-proficient viewers spent shorter time fixating on captions than more proficient participants, suggesting that learners who perceived their level of proficiency as too low for comprehension simply did not make the effort to process captions. The use of subtitles in the learners' native language may be a way to achieve this.

As stated before, the majority of studies on comprehension concur that subtitles facilitate understanding of the content better than captions (Bianchi & Ciabattini, 2008; Birulés-Muntané & Soto-Faraco, 2016; Latifi et al., 2011; Lwo & Lin, 2012; Markham, Peter & McCarthy, 2001; Markham & Peter, 2003), which is not surprising, since reading the text in your native language logically facilitates understanding. Subtitles are

processed automatically and provide on-line translations (Sydorenko, 2010). Also, learners tend to be better at reading than listening, and they can benefit from seeing difficult content on their native language first (Markham et al., 2001). Subtitling, however, is normally discouraged within foreign language classroom settings, as it is believed that the availability of L1 text will stop students from listening to the foreign language, and that they will focus only on reading the L1 text and will not listen to the L2 audio (Danan, 2004; Steward & Pertusa, 2004). Mitterer and McQueen (2009) also argued that subtitles create lexical interference and that negatively affect speech perception – whereas captions aided it by indicating what words are being spoken, facilitating the aural-written connection. This would suggest that, while L1 subtitles may help viewers identify the meanings of unknown words, they hinder adaptation to the L2 sound (Vanderplank, 2016a).

The benefits of subtitles in the native language have been explained by Danan (2004). When learners are exposed to more than one language at a time together with a bimodal input situation, three independent systems interconnect (image, sound in the L2 and text in the L1), which “may lead to a better processing and recall because of the additive effects of both image and translation” (Danan, 2004: 72). Therefore, “it seems perfectly sensible to exploit it [the L1] when it is to our advantage” (Schmitt, 2008: 337), which seems particularly true for younger viewers. In a study with young learners including captions and subtitles, Bravo (2008) reported that the absence of the L1 required a greater effort and higher L2 fluency amongst her young participants when completing post-viewing tasks. Additionally, research on reading has observed that if learners are unable to follow the overall story, they do not pay attention to the precise meaning of words (Laufer, 2005).

While the consensus is that captions (as they provide more exposure to the foreign language) are more beneficial for language learning and vocabulary acquisition,

adequate comprehension of written and aural input is also crucial for acquiring new words incidentally (e.g., Lee & VanPatten, 2003). Adequate comprehension of the content is likely to result in higher vocabulary retention, as the initial understanding of the content may allow learners to better concentrate on the unknown L2 input (Baltova, 1999). In a study including various forms of captioning, Montero-Perez, Peters, Clarebout and Desmet (2014) reported low scores in meaning recall across experimental groups, and suggested that one of the reasons might have been that, since captions did not provide concrete information on word meaning, learners had to construct meaning through inferences. Because of the real-time nature of listening, learners have little time to guess the meaning from the context; at the same time, focusing on the specific meaning of one word makes it more difficult for learners to listen to what is coming next, hindering coherent understanding (Goh, 2000).

Altogether, the above studies suggest that subtitles are generally more useful for comprehension than captions, especially for beginner learners. Subtitles provide on-line translations and allow understanding of the content regardless of the learners' language skills. Captions, on the other hand, can help learners with written and aural form recognition and with making form-meaning connections, but learners' bottom-up processing may be negatively affected if their vocabulary knowledge is limited. In the case of young learners, the few existing research findings show the important role played by proficiency but they are inconclusive as regards the effects of captions and subtitles on comprehension and vocabulary learning. Inconsistent results may also be due to developmental differences. In the following two sections, findings from comparative studies on viewing comprehension and vocabulary learning through subtitled and captioned audio-visual materials will be discussed.

## 2.4. Viewing comprehension

### 2.4.1. Studies on viewing comprehension

Previous research examining comprehension of audio-visual input has consistently shown the positive effects of captioning (or keyword captioning) over non-captioning (BavaHarji, et al., 2014; Chung, 1999; Garza, 1991; Gass, Winke, Isbell & Ahn, 2019; Guillory, 1998; Huang & Eskey, 1999; Li, 2014; Montero-Perez, Peters & Desmet, 2014; Montero-Perez, Van Den Noortgate & Desmet, 2013; Park, 2004; Rodgers, 2013; Rodgers & Webb, 2017; Winke et al., 2010). Although some studies have suggested that the presence of on-screen text might impose a cognitive burden for beginner learners (Taylor, 2005) or might be distracting for more advanced students (Baristow & Lavaur, 2011), the cognitive load that it adds seems not to be so detrimental as to hinder comprehension (Birulés-Muntané & Soto-Faraco, 2016). It has also been observed that aural and verbal textual information are processed in parallel, which would imply that the presence of text does not hinder the processing of the audio (Danan, 2004; d'Ydewalle & Gielen, 1992).

Virtually all studies comparing viewing comprehension with captions versus non-captions report a positive effect from captions over no text – although the differences in the percentage of comprehension between captions and non-captions conditions varies widely from study to study. Table 2.1 includes several exemplifying studies in this area, including a variety in input exposure, as well as differences in item number and format<sup>4</sup>. As can be observed, differences between conditions range from 2.6% to 46.8%, with the

---

<sup>4</sup> Studies with less generalizable or comparable results, or that have report incomplete statistical data have been not included in the table (e.g. Baltova, 1999; Etamadi, 2012; Guichon & McLoran, 2008; Garza, 1991; Matielo, de Oliveira & Baretta, 2017).

captions group consistently outperforming the no-text condition. Differences are significant in the majority of studies, with only few studies reporting overall non-significant differences (Montero-Perez, Peters, Clarebout & Desmet, 2014; Rodgers, 2013; Rodgers & Webb, 2017).

Table 2.1. Mean difference in comprehension percentages between captioning and non-captioning conditions across viewing comprehension studies

Study	Input and testing	Captions	No-text	Diff <sup>a</sup>
Birulés-Muntaner & Soto-Faraco, 2016	60-min full episode 8 items format n/a	79.4%	59.4%	20%* <sup>b</sup>
Chung, 1999	video fragment ( <i>length n/a</i> ) 10 items MC (4 options, L1)	76.6%	66.9%	9.7%*
Guillory, 1998	2 video fragments ( <i>length n/a</i> ) 2 tests, 7 items each short answer (L1)	71.8%	52%	19.8%*
Hayati & Mohmedi, 2011	six 5-min fragments 6 tests x 10 items each MC (L1)	66.3%	41%	25.3%*
Huang & Eskey, 1999	14-min episode (viewed twice) 16 items MC (3 options, L1)	67.9%	47.9%	20%*
Latifi, Mobalegh & Mohammadi, 2011	fifteen 2-min fragments 15 tests x 10 items each MC	66.3%	52.5%	13.8%*
Li, 2014	20-min episode 18 items MC (4 options, L1)	84.7%	37.9%	46.8%*
Markham, Peter & MaCarthy, 2001	7-min episode 10 items MC (L1)	56.7%	46.1%	10.6%*
Markham & Peter, 2003	7-min episode 20 items MC (4 options, L2)	50.6%	39%	11.6% <sup>c</sup>
Montero-Perez, Peters & Desmet, 2014	3 clips (8, 6, 2 min. respectively) 43 items total mixture of formats (L1)	59.6%	54.2%	5.4%*
Montero-Perez, Peters, Clarebout & Desmet, 2014	3 clips (2, 4, 3 min. respectively) 41 items total OE, MC and COMB (L1)	71.8%	68.1%	3.7%
Rodgers, 2013 Rodgers & Webb, 2017	ten 43-min episodes 10 tests x mean 74.2 each, 742 total MC (3 options), TF and SEC (L1)	66.5%	63.9%	2.6% <sup>d</sup>

**Types of format:** MC = Multiple-choice; OE = Open-ended; COMB = combination (e.g., match sentence and picture); TF = True-False; SEC = sequencing (e.g., order events chronologically)

<sup>a</sup> Difference between the caption group and the non-caption group (i.e. the score in the caption group was XX% higher than the non-caption group). To facilitate comparison, the difference is displayed in percentage – if not available, the percentage has been calculated based on the mean score and the number of items. Also, only results from the captions and non-captions group are reported, although some of the studies mentioned include other experimental conditions such as key-word captioning, L1 subtitles – which also outperform the no-text conditions, or advanced organizers.

<sup>b</sup> \* indicates difference between groups was statistically significant

<sup>c</sup> No statistical data reported

<sup>d</sup> Rodgers (2013) and Rodgers & Webb (2017) found that the mean difference from the 10 episodes was not significant. It reached statistical significance, however, in three of them (with a mean difference of 4.5% in those three).

Studies comparing subtitling versus captioning have generally found – as mentioned before – that the availability of subtitles in the learners’ native language facilitates comprehension better than captions, and consequently better than no-text (Bianchi & Ciabattoni, 2008; Birulés-Muntané & Soto-Faraco, 2016; Latifi et al., 2011; Lwo & Lin, 2012; Markham, Peter & McCarthy, 2001; Markham & Peter, 2003). Scant exceptions have favoured captions (Hayati & Mohmedi, 2011) and some studies have reported inconclusive results (e.g., Baristow & Lavour, 2011; Guichon & McLoran, 2008; Matiolo, Oliveira & Baretta, 2017; Vulchanova et al., 2015). Table 2.2 below provides a summary of the studies comparing viewing comprehension rates with subtitles and captions, including details of the studies methodological features (participants, input, conditions, testing instruments), and overview of the main results. Studies including young learners are marked with the symbol “★”, and studies including vocabulary learning measures alongside comprehension are marked with the symbol “❖”. In order to facilitate the comparison across studies, comprehension rates have been transformed into percentages when they were not available in the original study. When there is more than one study, only the one relevant for the discussion is presented – the study number is indicated in the first column. If more than one study is relevant, they appear in different rows to ease comparison of results.

Table 2.2. Studies comparing the effects of subtitles and captions on viewing comprehension from audio-visual input

Study	Participants	L2 proficiency	L1	L2	Input characteristics	Viewing conditions	Testing	Results
Bairstow & Lavaur (2011) STUDY 2 ★  Lavaur & Bairstow (2011) ★	90 secondary school students (aged 15-18)	Three proficiency groups: Beginner Intermediate Advanced  (translation task and auto-evaluation task)	Fr	En	8-9 min clip of a movie ( <i>North by Northwest</i> )  <b>Total viewing time:</b> 8-9 min  <b>V sessions</b> <sup>a</sup> : 1 (individual laptops)	Within each proficiency group: 1. subtitles 2. captions 3. no-text <sup>b</sup>	A 42-item comprehension test, including: (a) 21 items regarding visual elements (b) 21 items regarding dialogue elements	<i>Beginners:</i> Visual: no-text (65.2%) > captions (54.3%) > subtitles (44.3%) Dialogue: subtitles (69%) > captions (32.8%) > no-text (22.8%)  <i>Intermediate:</i> Visual: captions (52.4%) > subtitles (50.5%) > no-text (47.6%) Dialogue: subtitles (58.6%) > captions (55.7%) = no-text (54.8%)  <i>Advanced:</i> Visual: no-text (66.2%) > captions (57.1%) > subtitles (50.9%) Dialogue: no-text (85.7%) > captions (74.8%) = subtitles (74.3%)  ❑ Statistical data insufficiently reported
Bianchi & Ciabattoni (2008) ❖	85 university students (aged 18-45)	Three proficiency groups: Beginner Intermediate Advanced	It	En	Selected excerpts from 2 films ( <i>Harry Potter</i> and <i>Fantasia</i> )  ❑ Participants could re-watch clips <i>after</i> completing the task and review answers up to two times  <b>Total viewing time:</b> n/a  <b>V sessions:</b> 1 (individual computers with headphones)	Within each proficiency group: 1. subtitles 2. captions 3. control (= no-text)	Multiple-choice questions on comprehension of content, vocabulary and use of lexicogrammatical expressions (number of items n/a)	<i>All proficiency groups</i> Subtitles > captions > no-text (for the <i>intermediate</i> group, there was no difference between the captions and no-text conditions)  ❑ Statistical data insufficiently reported. Percentages on vocabulary gains are the mean between the two clips



Birulés-Muntaner & Soto-Faraco (2016) ❖	60 university students (aged 21-28)	B2 intermediate	Ca Es It	En	1 full episode (68 minutes) of a British TV drama ( <i>Downtown Abby</i> )  <b>Total viewing time:</b> 60 min  <b>V sessions:</b> 1 (setting n/a)	1. subtitles 2. captions 3. no-text	<b>Listening</b> Pre/post-test design An 24-item fill-in-the-gaps listening test  <b>Comprehension:</b> An 8-item comprehension test (format n/a)	<b>Listening</b> Captions (16.88%) > no-text (7.08%) > subtitles (0%)  Significant improvement with captions and no-text  <b>Comprehension</b> Subtitles (93.13%) > captions (79.38%) > no-text (59.38%)  Significant difference between subtitles and captions (.015), captions and no-text (>.001) and subtitles and no-text (>.001)
Bravo (2008) STUDY 2 ★❖	77 Grade 9 students (aged 13-14)	A2/B1	Pr	En	10 episodes (15 minutes each) of an American TV series comedy ( <i>The Fresh Prince of Bel-Air</i> )  <b>Total viewing time:</b> 150 minutes  <b>V sessions:</b> 10 (classroom)	1. subtitles 2. captions	<b>Comprehension + Vocabulary</b> After each episode: A 20-item multiple-choice test, including: - 10 items on content comprehension - 10 items on lexical expressions	<b>Comprehension + Vocabulary</b> ☑ <i>Results from vocabulary and comprehension post-viewing tests are reported together</i>  From episode 1 to 3: captions > subtitles From episode 4 to 10: subtitles > captions (only significant in two episodes)
Guichon & McLoran (2008)	40 university students	n/a  <i>“students were relatively inexperienced L2 listeners”</i>	Fr	En	1 BBC news report (3 min) about a struggling family  ☑ Note-taking allowed during viewings  <b>Total viewing time:</b> 3 min x 2 viewings = 6 min  <b>V sessions:</b> 1 (setting n/a)	1. subtitles 2. captions 3. no-text 4. audio-only	A 20-minute written summary (assessing number of central semantic units)	Captions (30.2%) > subtitles (29.7%) > no-text (25.1%) > audio-only (19.7%)  ☑ No statistical analyses were performed due to the small sample size

Hayati & Mohmedi (2011)	90 university students (mean age 22)	intermediate (standard English Language Proficiency Test)	Fa	En	6 excerpts (5 min each) of a documentary on natural disasters ( <i>Wild Weather</i> )  <b>Total viewing time:</b> 34 min  <b>V sessions:</b> 6 (language laboratory)	1. subtitles 2. captions 3. no-text	<b>Comprehension</b> A 10-item multiple-choice test after each video (6 tests in total)  <b>Perceptions</b> Written comments on their viewpoints about the effect of different types of on-screen text on films comprehension.	<b>Comprehension</b> Captions (66.37%) > subtitles (50.87%) > no-text (41.03%)  Significant difference between captions and subtitles (.000), captions and no-text (.000), subtitles and no-text (.011)  <b>Perceptions</b> Captions enhanced ability to notice, comprehend, spell and recall new words; subtitles facilitated recognition by providing the meaning.
Lafiti, Mobalegh & Mohammadi (2011)	36 students from a summer course (aged 17-33)	intermediate (in-house test)	Fa	En	15 clips (2 min each) of a cartoons film ( <i>Alvin and the Chipmunks</i> )  ☑ one clip x session, two screenings x clip  <b>Total viewing time:</b> 30 min x 2 screenings = 60 min  <b>V sessions:</b> 15 (classroom)	1. subtitles 2. captions 3. no-text	<b>Listening</b> IELTS listening test, as a pre-and post-test for listening skills  <b>Comprehension</b> A 10-item multiple-choice comprehension test per each clip	<b>Listening</b> No-text (5.74) > captions (5.63) > subtitles (5.54)  No significant differences in pre-test (.957), and significant in post-test (.017), but only between no-text and subtitles groups  <b>Comprehension</b> Subtitles (71.67%) > captions (63.33%) > no-text (52.50%)  Significant difference between subtitles and no-text (.000), captions and no-text (.011), and marginal between subtitles and captions (.061)
Markham, Peter & McCarthy (2001)	169 university students	intermediate (in-house general proficiency test)	En	Es	1 short clip from documentary (about Apollo 13 NASA)  <b>Total viewing time:</b> 7 min  <b>V sessions:</b> 1 (classroom)	1. subtitles 2. captions 3. no-text	A 10-minute written summary (assessing number of idea units)  A 10-item multiple-choice test	<i>Mean idea units</i> Subtitles (12.40) > captions (10.97) > no-text (8.47)  <i>Multiple-choice</i> Subtitles (82.4%) > captions (56.7%) > no-text (46.1%)  ☑ Statistical data insufficiently reported. Unclear whether difference reach significance between the captions and subtitles groups.

Markham & Peter (2003)	213 university students	intermediate (in-house general proficiency test)	En	Es	1 short clip from documentary (about Apollo 13 NASA)  <b>Total viewing time:</b> 7 min  <b>V sessions:</b> 1 (classroom)	1. subtitles 2. captions 3. no-text	A 20-item multiple-choice listening comprehension test	Subtitles (66.65%) > captions (50.60%) > no-text (39.05%)  Participants' last course grade in Spanish proved to be a significant predictor of success ( $p < .01$ )  ☐ Statistical data insufficiently reported
Matielo, Oliveira & Baretta (2017)	36 participants (aged 18-60; mean age 22)	Intermediate (in-house placement test)	Pt	En	1 full episode (20 min) of an American TV sitcom ( <i>The Big Bang Theory</i> )  <b>Total viewing time:</b> 20 min  <b>V sessions:</b> 1 (classroom)	1. subtitles 2. captions 3. control (no-text)	For general comprehension: 3 open question test ( <i>why...?</i> )  For specific comprehension: 5 true/false statements	<i>General</i> Subtitles (89%) > captions (75%) > no-text (58.33%) Significant difference only between subtitles and no-text condition (.009)  <i>Specific</i> Subtitles (93.4%) > captions (86.6%) > no-text (76.6%)  Significant difference only between subtitles and no-text condition (.014)  Effect of availability of text in both cases
Vulchanova, Aurstad, Kvitnes & Eshuis (2015) ★❖	114 secondary school students (aged 16-17)	n/a  "17-year-old group was assumed to have a higher level of proficiency in English"  (Cambridge Essential Grammar in Use Level Test + Vocabulary size test)	No	En	1 full episode (20 min) of an American TV series cartoon ( <i>Family Guy</i> )  <b>Total viewing time:</b> 20 min  <b>V sessions:</b> 1 (classroom)	1. subtitles 2. captions 3. no-text	A multiple-choice content comprehension test  + background questionnaire on L2 self-assessed skills	<i>17-year-olds</i> No statistical differences between captions and subtitles, but both significantly better than no-text (.000)  Other predictors: L2 self-assessed skills, grammar competence, being familiar with the target TV series  <i>16-year-olds</i> Captions > subtitles > no-text  Other predictors: vocabulary size, L2 self-assessed skills, playing computer games  ☐ Statistical data insufficiently reported

Yoshino, Kano & Akahori (2000) ❖	32 university students (3 <sup>rd</sup> /4 <sup>th</sup> year)  104 junior college students [vocational training] (1 <sup>st</sup> year)	n/a	Ja	En	4 music video clips (33 to 38 seconds each, with four different speech rates)  📺 music video  <b>Total viewing time:</b> 2.5 min x 2 viewings ≈ 5 min  <b>V sessions:</b> 1 (classroom)	1. subtitles 2. captions 3. no-text 4. audio-only (no video, no text)	Test on accuracy of recalled information (i.e. percentage of recalled clauses)	<i>University:</i> Subtitles (51.38%) > captions (39.92%) > no-text (15.14%) > audio-only (8.36%)  <i>Junior college:</i> Subtitles (39.75%) > captions (31.40%) > audio-only (1.51%) > no-text (1.24%)  Older students perform significantly better in all conditions
----------------------------------	--	-----	----	----	--	--	--	--

★ Study on young learners / ❖ Study including both comprehension and vocabulary tests / 📺 Noticeable methodological difference or unclear statistical results – comparison with other studies should be done with caution

Language codes: Ar = Arabic; Ber = Berber; Ca = Catalan; En = English; Es = Spanish; Fa = Persian; Fr = French; It = Italian; Ja = Japanese; Ku = Kurdish; NI = Dutch; No = Norwegian; Prs = Dari (Afghan Persian); Pt = Portuguese; Ro = Romanian / Moldovan; Ru = Russian; Zh = Chinese

<sup>a</sup> V sessions: number of viewing sessions (not including testing-only sessions; pre- and post-tests are often conducted on separate dates)

<sup>b</sup> The condition *no-text* and *control* are often the same, but the names given by the study appear in the Table (control groups are theorized differently in each study and the no-text can be seen as either a control group or another experimental group, depending on the research questions).

As can be observed, in contrast with the generally wider gap in comprehension found between text and no-text conditions, differences between captions and subtitles tend to be smaller, although still significant in many studies, as summarized in Table 2.3. Note that comparison between studies should be done, again, with caution; the table provides an overall picture of the general tendency between experimental conditions (through mean comprehension percentages), but does not take into account key factors such the proficiency level of participants or the type of assessment of comprehension.

*Table 2.3. Comparison of the percentage in comprehension with subtitles and captions across studies*

<b>Study</b>	<b>Subtitles [L1]</b>	<b>Captions [L2]</b>	<b>No-text</b>
Birulés-Muntaner & Soto-Faraco, 2016	93.1%	79.4%	59.4%
Bairstow & Lavour, 2011 Lavour & Bairstow, 2011	67.3%	49%	54.4%
Guichon & McLoran, 2008	29.7%	30.2%	25.1%
Hayati & Mohmedi, 2011	50.8%	66.3%	41%
Latifi, Mobalegh & Mohammadi, 2011	71.6%	66.3%	52.5%
Matielo, Oliveira & Baretta, 2017	91.2%	80.8%	67.5%
Yoshino, Kano & Akahori, 2000	45.5%	36.7%	8.2%

#### *2.4.1.1. Young learners*

Few studies have compared the effects of captions and subtitles on comprehension with young learners, reporting controversial results contingent upon proficiency and age. Baristow and Lavour (2011) and Lavour and Baristow (2011) investigated the comprehension of a 9-minute clip by secondary school students (aged 15-18) at three levels of proficiency, comparing subtitles, captions and non-captions. They found that, for the more advanced learners, having on-screen text was distracting, and that the non-captions group outperformed the other two – who performed similarly. However, for beginner learners it was found that on-screen text had a facilitating effect, and that the

subtitles group significantly outperformed the others. They also found that visual and dialogue information was recalled differently depending on viewers' proficiency level and language of the subtitles. Lwo and Lin (2012) studied the differential effects of captions and subtitles using a multimedia animated reading tool with Grade 8 learners and found that comprehension also depended on learners' proficiency and that differences between the language groups was not significant. They also reported that less proficient students were not overloaded with too much information, and that it seemed they could select what they needed from the input available. While the study did not include video input, results indicated that learners at age could access the on-screen text in either language – even if a different degrees. Vulchanova et al. (2015) looked at the comprehension of a 22-minute episode by 16- and 17-year-olds. For the older group, they found no significant differences between language groups, but for the younger group those in the caption condition performed better, although the authors note that for that age group the most significant predictor was vocabulary size rather than language of the subtitles. Altogether, the above studies suggest that, while there seems to be a quite general consensus that subtitles are more useful for comprehension than captions (and specially for beginners), there are still aspects of this relationship that are not known, such as the potentially different effects of age and proficiency, or the experimental conditions of the study (e.g., type of information that is being asked for).

#### **2.4.2. Comprehension in extensive viewing. Longitudinal studies**

Rodgers and Webb (2011) found that related television programmes are likely to contain fewer word families than unrelated programmes, and that word families from the 4,000–14,000 levels were more likely to reoccur in a complete season of a television programme than in a random sample of television programmes. This suggests that the

more episodes you watch from the same TV series, the more potential to learn from them (Webb & Rodgers 2009a). However, as seen above, most studies have used short clips, segments of films, or educational videos, which are largely unrelated, and not fully representative of what a viewer would normally choose to watch (Rodgers, 2013). While there have been studies that have used full movies, they were shown divided in sections and had several viewing sessions (Naghizade & Darabi, 2015; Steward & Pertusa, 2004; Zarei & Saddeghi, 2011), or used only fragments of the original input (Gorjian, 2014; Hayati & Mohmedi, 2011), which does not seem a very natural way of watching this type of input. In fact, only a few studies have used longer, uninterrupted, authentic input, such as full TV episodes, documentaries or movies (e.g., Peters & Webb, 2018).

Longitudinal classroom-based studies using several, full-length TV episodes to investigate comprehension (or vocabulary learning) are scarce as well, but the number is rising. Zarei (2008) used nine episodes of a British comedy to assess vocabulary acquisition and comprehension. Bravo (2008) used 10 episodes of an American comedy to investigate learning of lexical expressions and content comprehension with beginner learners. Rodgers (2013) also investigated incidental vocabulary learning and comprehension through the viewing of 10 episodes of a TV series, and the effects on frequency and range of occurrence. BavaHarji et al. (2014) used 30 episodes of a TV series to examine the effects of captioned instructional videos on EFL learners' vocabulary acquisition and language proficiency. Frumuselu et al. (2015) studied the acquisition of informal and conversational speech through 13 episodes of a subtitled TV series. Chen, Liu and Todd (2018) explored spoken vocabulary acquisition through 10 episodes of an animated television series. Gesa (2019) investigated vocabulary learning and comprehension at different age groups (Grade 6, Grade 10 and university students) through 24 episodes of a TV series, and with the additional support of pre-viewing activities. The length of each intervention and the spacing between the viewings (as well

as the number of viewings) varies largely between studies, but these longitudinal data may be able to reproduce more accurately the situation that learners encounter in real-life (e.g., accumulation of background knowledge, repeated exposure to vocabulary in the genre).

From the aforementioned studies, only a few address the issue of whether an extended exposure to audio-visual material could have a training effect, and whether an improvement could be seen over time. Bravo (2008) found that participants' performance on the post-viewing tasks (which included questions regarding content comprehension and lexical expressions) changed over time, with a marked improvement after the fourth episode – regardless of the language of the on-screen text. Rodgers (2013) found that the perceived comprehension of the episodes increased as more episodes were viewed, quickly improving after the initial episode. Comprehension scores corroborated students' perception, showing a significant increase of 8% in comprehension from the first to the last episode. On the other hand, Gesa (2019), in a study that included 24 viewing sessions from two different TV series, did not find an improvement in comprehension scores. In general, however, findings indicate that the more episodes you watch from a TV series, the easier it gets to understand its content, and therefore the more potential to learn from it.

### **2.4.3. Lexical coverage**

Lexical coverage – that is, the percentage of known words in the input – provides an indication of the vocabulary size needed for adequate comprehension of a specific text, together with the vocabulary load that it represents (Webb, 2011; Webb & Rodgers, 2009b). The higher the lexical coverage, the easier it might be for learners to understand the content (Webb, 2011). The lexical coverage of the episodes also plays a prominent



role in comprehension, beyond the proficiency level of the learners itself. A very recent state-of-the-art article by Nurmukhamedov and Webb (2019) provides a valuable overview of the studies on lexical coverage, and the major areas of lexical coverage research.

Research on reading and listening – and more recently on TV and film viewing – has extensively shown that vocabulary knowledge is a strong predictor of content comprehension (e.g., Hu & Nation, 2000; Laufer & Ravenhorst-Kalovski, 2010; Rodgers, 2013; Van Zeeland & Schmitt, 2012; Webb & Rodgers, 2009b), although disagreement exists on the percentage of lexical coverage needed for adequate understanding of the written, aural or audio-visual input. Research on extensive reading suggests that learners need up to 95% coverage for minimal comprehension and 98% coverage for optimal comprehension (e.g., Hu & Nation, 2000; Laufer, 1989; Laufer & Ravenhorst-Kalovski, 2010), while research on informal listening proposes a less conservative figure, suggesting that a coverage of 90-95% might be enough to understand everyday conversations (e.g., Bonk, 2000; Van Zeeland & Schmitt, 2012). For viewing comprehension, it has also been suggested that 95% might be enough, because of the additional support provided by images (Rodgers & Webb, 2011). Rodgers (2013) investigated comprehension of a series of TV episodes and found that comprehension improved with increased lexical coverage, although not for all episodes, which indicates the need to take into account differences on episodes' lexical coverage when assessing comprehension across different videos. It is important to bear in mind that the coverage needed for comprehension can vary depending on the type of discourse and the degree of comprehension required (Rodgers & Webb, 2009b). The percentage needed for “adequate” or “reasonable” comprehension is not a clearly defined term, and it can vary across studies and – as pointed out above – input modes (i.e. reading, listening or viewing). In addition, it has been argued that learners may be able to operate at lower

levels of coverage in real-world situations, where more information is provided by the conversational context (Bonk, 2000).

While the abovementioned studies investigated the percentage of lexical coverage needed to reach an “adequate” level of comprehension, research in this area has also looked at the number of words necessary to reach that percentage. For reading comprehension, first studies suggested that knowing 3,000 word families was the minimum to understand unsimplified texts (Laufer, 1992), while others estimated a minimum of 5,000 words to read short novels (Hirsh & Nation, 1992). In order to understand spoken English, it is generally accepted that language learners need to know around 2,000–3,000 word families to reach 95% coverage, and 6,000–7,000 word families to reach 98% coverage (Nation, 2006). Adolphs and Schmitt (2003) found that 2,000 word families might be insufficient, and suggest 3,000 word families as a more appropriate goal to understand informal everyday conversations.

For viewing comprehension, the amount of words necessary to achieve the threshold appears to be a bit lower. Rodgers and Webb (2009b) analysed the scripts of 318 American and British movies and found that the knowledge of the most frequent 3,000 word families – plus proper nouns (e.g., Eddie, Orlando) and interjections (e.g., ah, wow) – already provided 95% coverage. Rodgers and Webb (2008a) analysed the scripts of 88 TV programmes and also found that 95% coverage could be reached with the knowledge of 3,000 word families. The authors also found that lexical coverage varied across genres, and that the coverage needed within a genre might be lower due to the repetition of vocabulary if learners keep on viewing programmes within the same genre – as seen above – and learn its most common vocabulary. Rodgers (2019) also found differences in the lexical coverage across different types of games. Nurmukhamedov’s (2017) study provides further evidence of the differences in vocabulary demands from audio-visual materials. Nurmukhamedov analysed the transcripts of 400 TED Talks

presentations, and found that knowledge of the 4,000 words families (plus proper nouns and marginal words) provided 95% coverage, and 8,000 word families were needed to achieve 98% coverage. The results suggest that, for TED Talks, the coverage necessary for adequate comprehension resembles more that of a written text than a TV series. In spite of the variance across studies, findings from the field provide cumulative evidence of the importance of lexical coverage and the need to take this factor into account.

#### **2.4.4. Testing instruments**

Besides factors related to the viewer (e.g., learners' proficiency) and the input (e.g., lexical coverage), research in reading and listening comprehension have revealed that differences in testing yield varying degrees of difficulty for test-takers, with significant effects on comprehension scores depending on input materials, question format and language used – especially for beginner learners (e.g., Shohamy, 1984). In order to design appropriate tasks, we need to establish first what is assessed (construct validity) (Vandergrift, 2007). Buck (2001) proposed a flexible, baseline definition of the listening construct adequate for L2 classroom assessment that describes listening comprehension 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 context of the passage.” (Buck, 2001: 114)

This definition seems to be appropriate for comprehension through TV input too, considering that even if the addition of visual support may facilitate information processing, we still assess viewers' ability to understand what is being said. Related to this, Wagner (2002) investigated construct validity of a video-based test, and found

evidence for the validation of a two-factor model based on the ability of processing explicit information and implicit information in aural input, instead of the hypothesized top-down and bottom-up factors. This concurred with previous research that had already called attention to these two main types of questions, with numerous variations regarding their nomenclature and possible sub-typologies (Alptekin & Erçetin, 2010; Buck, 2001; Davey & McBride, 1986; Pearson & Johnson, 1978; Rodgers, 2013; Shohamy & Inbar, 1991). Most commonly, textually explicit or literal questions refer to items that ask for information explicitly stated in the text (information that could be underlined), regarding details or trivial information, and they normally involve bottom-up processing. Textually implicit or inferential questions, on the other hand, ask for information that is found by integrating different pieces of information and making inferences, involving top-down processing. This type of questions can include going beyond the text to understand the central gist or idea or synthesizing information to draw conclusions. Although still in need of more research, studies including different question types indicate that item type has an effect on comprehension scores and that the presence of on-screen text – or the absence of it – interacts with item type (e.g., Rodgers, 2018b; Shohamy & Inbar, 1991). Item format – that is, how questions are presented – also deserves attention. Response format can significantly affect comprehension scores, as shown by, for example, Cheng's (2004) study, where learners completing multiple-choice items outperformed respondents of open-ended items.

## 2.5. Vocabulary learning

### 2.5.1. Studies on incidental vocabulary learning

Another area of interest in research on language learning through viewing has been its potential for vocabulary learning. Studies on vocabulary acquisition through audio-visual input have generally been focused on incidental learning. The term *incidental learning* has various definitions (de Vos, Schriefers, Nivard & Lemhöfer, 2018). A first definition can be made from the perspective of the learner, when learning takes place “without intention, while doing something else” (Ortega, 2014: 94). A second definition applies in the context of research and is dependent on the announcement of an upcoming test, where incidental learning occurs when there is no expectation of being tested afterwards (Hulstijn, 2003). A third definition is based on the nature of the activity itself, when learning occurs as a by-product of another activity. The bulk of research on vocabulary acquisition through audio-visual input has focused on incidental vocabulary learning understood in this last sense, where learning occurs as a by-product of another meaning-focused activity (i.e. watching a video for its information content). A growing number of studies in the area consistently indicate that incidental vocabulary acquisition does occur through viewing short clips, full movies (e.g., Peters & Webb, 2018), and TV series (e.g., Rodgers, 2013), although the comparison of vocabulary gains is difficult due to differences in testing format and number of items.

As seen in previous sections, the general consensus is that viewing with captions provides more exposure to the target language, thus being more beneficial for language learning and vocabulary acquisition (e.g., Danan, 2004; Vanderplank, 2010; Winke et al., 2010). Indeed, the majority of comparative studies have found that captions have more positive effects on vocabulary learning than subtitles (Birulés-Muntaner & Soto-Faraco, 2016; Frumuselu et al., 2015; Matielo, Collet & D’Ely 2013; Naghizadeh & Darabi, 2015;

Zarei, 2008; Zarei & Rashvand, 2011). Other studies, per contra, show that more benefits are derived from subtitling – especially for low proficiency learners (Bianchi & Ciabattini, 2008)–, while some others report inconclusive results, with small or non-significant differences between captioning and subtitling (Bisson et al., 2014; Bravo, 2008; Steward & Pertusa, 2004). Table 2.4 shows a summary of comparative studies on vocabulary learning, including information on the treatment, the tests used and a summary of the results. Similar to the Table 2.2, studies including young learners are marked with the symbol “★”, and studies including vocabulary learning measures alongside comprehension are marked with the symbol “❖”. Again, in order to facilitate the comparison across studies, vocabulary gains have been transformed – if enough information was provided – into percentages.

Table 2.4. Studies comparing the effects of subtitles and captions on vocabulary learning through viewing audio-visual input

Study	Participants	L2 proficiency	L1	L2	Treatment	Conditions	Testing	Conclusions
Aloqaili (2014) ★	48 secondary school students (aged 16-17)	A1 beginners  (online OPT test)	Ar	En	7 min clip of a movie ( <i>An American Girl: Chrissa Stands Strong</i> )  ☑ Two screenings  <b>Total viewing time:</b> 7 min x 2 viewings = 14 min  <b>V sessions:</b> 1 (classroom)	1. subtitles 2. captions 3. dual 4. control (no-text)	Pre-/post-test design (a) 10-item Vocabulary Knowledge Scale test	Captions (80.8%) > subtitles (79.2%) > dual (74.2%) > no-text (15%)  Difference only significant between the experimental groups and the control group, but not between the three experimental groups.  For recognition, subtitles > captions > dual For comprehension, subtitles = captions > dual For production, captions > dual > subtitles
Bianchi & Ciabattoni (2008) ❖	85 university students (aged 18-45)	Three proficiency groups: Beginner Intermediate Advanced	It	En	Selected excerpts from 2 films ( <i>Harry Potter</i> and <i>Fantasia</i> )  ☑ Participants could re-watch clips <i>after</i> the task and review answers twice  <b>Total viewing time:</b> n/a  <b>V sessions:</b> 1 (individual PC, with headphones)	Within each proficiency group: 1. subtitles 2. captions 3. control (= no-text)	Pre/post-test design (a) multiple-choice cloze grammar test (b) general lexical test (c) target items vocabulary test (d) pragmatic use of lexico-grammatical phrases test	<i>Beginners</i> No-text > subtitles (63.5%) > captions (56.2%)  <i>Intermediate</i> Subtitles (72.3%) > captions (70.3%) = no-text  <i>Advanced</i> Subtitles (78.4%) = captions (78%) = no-text  ☑ Statistical data insufficiently reported. Percentages on vocabulary gains are the mean between the two clips
Birulés-Muntaner & Soto-Faraco (2016) ❖	60 university students (aged 21-28)	B2 intermediate	Ca Es It	En	1 full episode (68 minutes) of a British TV drama ( <i>Downton Abby</i> )  <b>Total viewing time:</b> 60 min  <b>V sessions:</b> 1	1. subtitles 2. captions 3. no-text	Pre/post-test design (a) 15-item definition matching test (target items were low frequency words, occurring 3-11 times)	No-text (8.67%) > captions (6%) > subtitles (-5%)  Significant improvement only in the no-text condition Inconclusive results for vocabulary.

Bisson, Van Heuten, Conklin & Tunney (2014)	54 participants (mean age 24, diverse educational background)  36 for the eye-tracking experiment	No prior knowledge of the L2 (self-reported)	En	NI	Excerpt of an animation film ( <i>SpongeBob Square pants</i> )  <b>Total viewing time:</b> 25 min  <b>V sessions:</b> 1 (eye-tracker)	1. subtitles 2. captions 3. reversed 4. control (no-text) 5. control (no video, only vocabulary test)	A 78-items aural meaning recognition post-test (i.e. select the correct translation over the distractor)	No differences amongst experimental groups (authors suggest the need for a more sensitive measure)  <b>Eye-tracking data</b> Text was read independently of the type of on-screen text  Reading becomes more regular when audio is in unknown foreign language  No differences between groups in gaze duration, number of fixations or skipped subtitles, with one exception: the captions group spent more time on the image area
Bravo (2008) STUDY 2 ★❖	77 Grade 9 students (aged 13-14)	A2/B1	Pr	En	10 episodes (15 minutes each) of an American TV series comedy ( <i>The Fresh Prince of Bel-Air</i> )  <b>Total viewing time:</b> 150 minutes  <b>V sessions:</b> 10 (classroom)	1. subtitles 2. captions	<b>Comprehension + Vocabulary</b> After each episode: A 20-item multiple-choice test, including: - 10 items on content comprehension - 10 items on lexical expressions  <b>Vocabulary</b> ("Consolidation tests") (b) Two 25-item multiple-choice test on lexical items [one after episode 5, one after episode 10] (c) 50-item multiple-choice test on lexical items [3 weeks after] (d) 50-item multiple-choice test on lexical items [3 months after]	<b>Comprehension + Vocabulary</b> ☒ <i>Results from vocabulary and comprehension post-viewing tests are reported together</i>  From episode 1 to 3: captions > subtitles From episode 4 to 10: subtitles > captions (only significant in two episodes)  <b>Vocabulary</b> ("Consolidation tests") CT1: captions (82.1%) > subtitles (78.3%) CT2: subtitles (80.8%) = captions (80.5%) CT3: subtitles (68.2%) > captions (66.1%) CT4: captions (65.7%) = subtitles (65.2%)  Inconclusive results



Frumuselu (2015)  Frumuselu, Maeyar, Donche, Gutiérrez (2015)	40 university students (BA English Studies) (aged 19-25)	A2 to C1 (not counter-balanced across groups)	Ca Es NI Ru Ro	En	13 full-episodes (25 min each) from an American sitcom ( <i>Friends</i> )  <b>Total viewing time:</b> 325 min  <b>V sessions:</b> 13 within 7 weeks (classroom)	1. subtitles 2. captions	Pre-/post-test design (a) 30-item test on “colloquial and informal expressions” [slang], including: - 15 multiple-choice meaning recognition items - 15 meaning recall items (translation task with context provided)	Captions (48.93%) > subtitles (36.5%)  Significant difference between language groups (.010)  <input checked="" type="checkbox"/> Heterogeneous proficiency level within groups, although author reports that results were “not dependent on proficiency”
Gorjian (2014)	90 university students (BA English translation) (aged 18-32)	intermediate (TOEFL)	Fa	En	8 movie fragments (10 minutes each) from 3 animation movies ( <i>The Ant Bully, The Emperor's new Groove, Kung Fu Panda 2</i> )  <b>Total viewing time:</b> 80 min  <b>V sessions:</b> 8 within 4 weeks (classroom)	1. subtitles 2. captions 3. reversed	Pre-/Post-test design (a) 40-item multiple-choice (meaning recognition?) test	Reversed (17.37) > captions (9.89) > subtitles (5.63)  Difference is only significant between reversed and subtitles condition  <input checked="" type="checkbox"/> Sessions included an in-class discussion on the films' content. Statistical data reported is insufficient.
Mardani & Najmabadi (2016) ★	60 high school students (aged 16-17)	n/a	Fa	En	Fragments (10 min each) from 3 animated movies ( <i>Tangled, Big Hero 6, Mr. Peabody Sherman</i> )  <b>Total viewing time:</b> 100 min  <b>V sessions:</b> 10 (classroom)	1. subtitles 2. captions 3. reversed	Pre-/post-test: (a) 30-item multiple-choice meaning recognition test (target items were presented in context, which was different in the pre- and post-test)	Reversed (95.83%) > captions (91.33%) > subtitles (89.83%)  Differences are only significant between reversed and captions (.016), and reversed and subtitles (.002)  <input checked="" type="checkbox"/> Sessions included an in-class discussion on the films' content.  Subtitles group's pre-test score was slightly lower than the others, although difference do not reach significance (.062)

Maritelo, Collet & D'Ely (2013)	27 participants (aged 16-48; mean age 22)	intermediate	Pt	En	1 full episode (20 min) of a TV sitcom ( <i>Big Bang Theory</i> )  <b>Total viewing time:</b> 20 min  <b>V sessions:</b> 1 (classroom)	1. subtitles 2. captions 3. control (=no text)	<b>Vocabulary</b> Pre-test: (a) 45-item (20 target items + 25 distractors) meaning recall test (translation task)  Post- & delayed post-test: [one week later] (b) 20-item meaning recall test (translation task)  <b>Perceptions</b> Questionnaire on viewing habits	<b>Vocabulary</b> <i>Gains from pre- to post-test</i> Captions (5%) > subtitles (0%) = control (0%)  <i>Gains from pre- to delayed post-test</i> Captions (5%) > subtitles (2.5%) = control (2.5%)  Not significant but positive effect of captions (but captions group scored higher in pre-test)  Authors suggest differences might be due to proficiency (which was not assessed)  <b>Perceptions</b> Subtitles > captions at home Teachers reported using more captions in class
Naghizade & Darabi (2015) ★	27 secondary school students (aged 15-17)	intermediate (OPT test)	Fa	En	6 sections (15 min) of a full movie ( <i>The impossible</i> )  ☑ Three screenings  <b>Total viewing time:</b> 90 min x 3 viewings = 180 min  <b>V sessions:</b> 6 (classroom)	1. subtitles 2. captions 3. no-text	Pre-test: (a) 96-item meaning recognition multiple-choice test  Post-test: (b) 40-item meaning recognition multiple-choice test (words known in the pre-test were excluded)	Captions (67.4%) > subtitles (57.7%) = no-text (57.7%)  Significant difference between captions and the other two conditions (.001)
Peters, Heynen & Puimège (2016) STUDY 1 ★	28 secondary school students (aged 17-18)	n/a  <i>Proficiency measure within the study: vocabulary size test</i>	NI	En	13-minute section of a documentary about eating insects  ☑ Two screenings  <b>Total viewing time:</b> 13 min x 2 viewings = 26 min  <b>V sessions:</b> 1 (classroom)	1. subtitles 2. captions	Pre-/post-test: (a) 50-item (39 target items + 11 non-words) spoken form recognition test (yes/no test) (b) 50-item ( <i>same</i> ) spoken meaning recall test (translation task)	<i>Form recognition</i> Captions (48.2%) > subtitles (32.4%) Difference was significant (.001) Effect of learners' VS and frequency  <i>Meaning recall</i> Subtitles (20.8%) > captions (19.3%) No significant differences Effect of learners' VS and frequency

Peters, Heynen & Puimège (2016) STUDY 2 ★	18 vocational school students (aged 17-20)	low-proficient to pre-intermediate  <i>Proficiency measure:</i> vocabulary size (VS) test	Nl Ber Prs Ru Ku	En	1 full episode (20 min) of a cartoon TV series ( <i>The Simpsons</i> )  <b>Total viewing time:</b> 20 min  <b>V sessions:</b> 1 (classroom)	1. subtitles 2. captions	Pre/post-tests: (a) 35-item (18 target items + 17 non-words) written form recall test (translation task) (b) 35-item ( <i>same</i> ) written form recognition test (yes/no) (c) 35-item ( <i>same</i> ) written meaning recognition multiple-choice test	<i>Form recall</i> Captions (21.5%) > subtitles (11.1%) Difference was significant (.016) Effect of learners' VS and frequency  <i>Form recognition</i> Captions (29.1%) > subtitles (25.5%) No significant difference Interaction between learners' VS and frequency  <i>Meaning recognition</i> Subtitles (31.8%) > captions (17%) No significant difference Interaction between learners' VS and frequency
Steward & Pertusa (2004)	95 university students	intermediate	En	Es	2 full movies, divided into 3 segments ( <i>Mujeres al borde de un ataque de nervios, Sexo por compasión</i> )  <b>Total viewing time:</b> 88 min + 109 min = 197 min  <b>V sessions:</b> 6 (classroom)	1. subtitles 2. captions	<b>Vocabulary</b> Pre-/post-test design (a) multiple-choice vocabulary recognition test   <b>Perceptions</b> Questionnaire on learners' experiences	<b>Vocabulary</b> Authors report that results are <i>inconclusive</i> , but there is a positive tendency of captions > subtitles  <input type="checkbox"/> Statistical data insufficiently reported. Number of items and results unknown  <b>Perceptions</b> High feeling of learning Although they are more challenging, students recognize the worth of making the effort to view films with captions
Vulchanova, Aurstad, Kvitnes & Eshuis (2015) ★❖	114 secondary school students (aged 16-17)	n/a  (Cambridge Essential Grammar in Use Level Test + Vocabulary size test)	No	En	1 full episode (20 min) of an American TV series cartoon ( <i>Family Guy</i> )  <b>Total viewing time:</b> 20 min  <b>V sessions:</b> 1 (classroom)	1. subtitles 2. captions 3. no-text	Delayed post-tests only [4 weeks later], including: (a) 30-item multiple-choice meaning recognition test (definition task) (b) 53-item (22 target items + 31 distractors) form recall test	Positive effect of subtitles in both tests, but modulated by age  <input type="checkbox"/> Statistical data insufficiently reported. No pre-test for vocabulary. Long lapse between exposure and post-tests.

Yoshino, Kano & Akahori (2000) ❖	32 university students (3 <sup>rd</sup> /4 <sup>th</sup> year)  104 junior college students [vocational training] (1 <sup>st</sup> year)	n/a	Ja	En	4 music video clips (33 to 38 seconds each, with four different speech rates)  ☑ music video  <b>Total viewing time:</b> 2.5 min x 2 viewings ≈ 5 min  <b>V sessions:</b> 1 (classroom)	1. subtitles 2. captions 3. no-text 4. audio-only (no video, no text)	Post-test only: (a) word recall ratio (number of recalled words / total words)	<b>Vocabulary</b> <i>University:</i> Captions (55.29%) > subtitles (25.97%) > no-text (14.90%) > audio-only (14.33%)  <i>Junior college:</i> Captions (51.91%) > subtitles (28.85%) > no-text (26.55%) > audio-only (20.33%)  Captions significantly better than the other conditions for both age groups (>.001), and no significant differences between age groups
Zarei (2008)	92 university students (aged 19 to 26; different BAs)	n/a Similar proficiency (Michigan test)	Fa	En	9 full episodes (30 min each) of a British comedy ( <i>Yes, Minister</i> )  ☑ Participants were asked to watch the films at home again  <b>Total viewing time:</b> 180 min (in class)  <b>V sessions:</b> 9 (classroom)	1. subtitles 2. captions 3. reversed	Pre-test: (a) Pre-test: 100-item multiple-choice  Post-test: (b) 40-item vocabulary recognition multiple-choice test (c) 40-item vocabulary recall fill-in-the-blanks test	<i>Recognition</i> Captions (82.58%) > subtitles (82%) > reversed (69.33%)  Only significant between captions and reversed (>.001), and subtitles and reversed (.001)  <i>Recall</i> Captions (67.25%) > subtitles (58.17%) > reversed (50.08%)  Significant between captions and reversed (>.001), subtitles and reversed (.039), and captions and subtitles (.017)
Zarei & Rashvand (2011)	120 participants (aged 19-32) conversational English course	Intermediate (TOEFL test)	Fa	En	Summarized version (70 min) of a movie ( <i>She's the man</i> )  <b>Total viewing time:</b> 70 min  <b>V sessions:</b> 1 (setting n/a)	1. subtitles verbatim 2. subtitles non-verbatim 3. captions verbatim 4. captions non-verbatim	<b>Vocabulary</b> Pre-test: (a) 100-item meaning recall test (provide an L1 synonym for the given word)  Post-tests: (b) TOEFL vocabulary subtest (c) 3-item meaning recognition* multiple-choice test	<b>Vocabulary</b> <i>Meaning recognition</i> Overall, captions > subtitles Within each language group, non-verbatim > verbatim  Difference is not significant regarding language, but it is regarding text content (.010)  <i>Form recall</i>

							(d) 3-item form recall fill-in-the-blanks test (first letter given) (e) written one-page summary [not analysed, used to ensure attention]	Captions > subtitles  Difference is not significant regarding text content, but it is regarding language (.006)  ☒ Statistical data insufficiently reported. Small number of items in relation to the length of the input.
Zarei & Saddeghi (2011)	120 university students (BA in translation)	n/a "same level of reading proficiency" (TOEFL test)	Fa	En	1 full documentary (50 minutes) divided in 3 segments of 17 min each ( <i>Planet Earth</i> )  <b>Total viewing time:</b> 50 min  <b>V sessions:</b> 3 (classroom)	1. subtitles synchronous 2. captions synchronous 3. subtitles asynchronous (reading script first) 4. captions asynchronous (reading script first)	<b>Vocabulary</b> Pre-test: (a) 217-item meaning recall test (translation of words in context)  Post-tests: (b) 30-item meaning recognition multiple-choice test (c) 30-item form recall fill-in-the-blanks test (L1 word given)	<b>Vocabulary</b> <i>Meaning recognition</i> captions > subtitles, but not significant  <i>Form recall</i> captions > subtitles, but not significant  ☒ Statistical data insufficiently reported. Different type of task at pre- and post-test.

★ Study on young learners / ❖ Study including both comprehension and vocabulary tests / ☒ Noticeable methodological difference or unclear statistical results – comparison with other studies should be done with caution

Language codes: Ar = Arabic; Ber = Berber; Ca = Catalan; En = English; Es = Spanish; Fa = Persian; Fr = French; It = Italian; Ja = Japanese; Ku = Kurdish; NI = Dutch; No = Norwegian; Prs = Dari (Afghan Persian); Pt = Portuguese; Ro = Romanian / Moldovan; Ru = Russian; Zh = Chinese

a V sessions: number of viewing sessions (not including testing-only sessions; pre- and post-tests are often conducted on separate dates)

b The condition no-text and control are often the same, but the names given by the study appear in the Table (control groups are theorized differently in each study and the no-text can be seen as either a control group or another experimental group, depending on the research questions).

As can be seen, for word form recognition and recall, studies concur that captions are more beneficial than subtitles (Peters, Heynen & Puimège, 2016; Yoshino et al., 2000; Zarei & Rashvand, 2011; Zarei & Saddeghi, 2011), which is not surprising since captions provide the spelling of the words in the target language. For word-meaning learning, however, results vary depending on the study. Most studies have reported inconclusive results for meaning recognition and recall, with non-significant differences between the language groups but with the captions conditions having a slight advantage over the subtitle ones (Birulés-Muntaner & Soto-Faraco, 2016; Lwo & Lin, 2012; Gorjian, 2014; Mardani & Najmabadi, 2016; Maritelo, Collet & D'Ely, 2013; Zarei, 2009; Zarei & Rashvand, 2011; Zarei & Saddeghi, 2011). Significant differences between the language groups were only found in Naghizade and Darabi's (2015) study for meaning recognition, and Zarei's (2009) study for meaning recognition and recall. Only one study found that the subtitles groups performed better than the captions groups in both meaning recognition and recall (Peters et al., 2016), although the difference was not significant in either case. This is somewhat surprising, since one would expect that having access to the L1 translation would be more useful. A reason might be that if the learner cannot make the connection between the aural and written form of the word, having the L1 on the screen might help overall content comprehension, but not understanding specific lexical items.

Mixed results in comparative studies on captioned and subtitled audio-visual input might be due to differences in methodology (test modality, length of exposure, target items) and the characteristics of participants, especially their proficiency level (Malone, 2008; Mohd Jelani & Boers, 2018). Participants' proficiency – when reported – ranges from beginner to advanced, sometimes even within the same sample (e.g., Frumuselu et al., 2015), which poses a significant problem when discussing results against other studies (Zarei & Rashvand, 2011). Importantly, it has been found that learners from

different proficiency levels show different responses to different on-screen text language within the same study, especially when learners are younger and less proficient (e.g., Lwo & Lin, 2012). Also, learners with larger vocabulary knowledge perform better than learners with smaller vocabularies (e.g., Horst, Cobb & Meara 1998; Peter & Webb 2018; Webb & Chang, 2015a), suggesting that more proficient participants will normally perform better.

#### *2.5.1.1. Young learners*

Proficiency level and age are usually closely linked. Most studies on vocabulary learning through audio-visual input and L2 learning have been conducted at university level or with adult language learners (e.g., Montero-Perez, Peters, Clarebout & Desmet, 2014; Sydorenko 2010; Winke et al., 2010). Although still scarce, the number of studies focusing on multimodal input with children and adolescents has increased in the last two decades, and research has demonstrated that watching subtitled or captioned television has positive effects also on their foreign language learning. Early studies observed that primary school children benefitted from subtitles in their L1 (Koolstra & Beentjes, 1999), and that even pre-schoolers could learn new L1 vocabulary through exposure to audio-visual input (Rice, Huston, Truglio & Wright, 1990). D'Ydewalle and Van Poel (1999) conducted a study with 8-12 year-olds testing incidental L2 learning through a 10-min still-motion movie comparing normal and reversed subtitles, and found that – even with a short exposure – participants in both conditions already showed small gains in vocabulary.

Several recent studies have focused on the comparison between young learners' viewing audio-visual material with or without subtitles or captions. Hsu et al. (2013) investigated the effect of subtitle mode on vocabulary acquisition with Grade 5

participants during a one-month experiment, comparing non-captioning, full-captioning and keyword-captioning. These researchers found that there were no differences between the two captioning groups and that both outperformed the non-captioning one in the meaning recognition test. Lekkai (2014) explored incidental vocabulary learning through a 15-minutes cartoon with Grade 4, 5 and 6 learners, with and without L1 subtitles. Learners in the subtitles group outperformed the no-text and control groups in both form recognition and meaning recognition, supporting the idea that, even at this young age, students can learn from subtitled video. Chen, Liu and Todd (2018) explored the effect of captioning (against non-captioning) on spoken vocabulary with Grade 8 learners, and found that the availability of captions significantly improved learners' recognition of form and form-meaning knowledge of novel L2 spoken vocabulary, especially for higher-proficiency learners. Gesa (2019) – as mentioned above – run a longitudinal classroom-based intervention including participants from Grade 6 and Grade 10, who watched several episodes of a TV series with subtitles and captions respectively. Although the study did not compare the two types of on-screen text (and the series were different), Gesa found that learners made progress in learning word forms and word meanings in both age groups, and that older learners outperformed the younger, less proficient ones.

Studies comparing the effects of on-screen text language (either L1 or L2) with young learners have also been recently conducted. Bravo (2008) compared the effects of watching captioned or subtitled episodes of a TV show on lexical expressions and comprehension scores for 13-14 year-olds. This researcher found similar results for both experimental groups, but also reported that the absence of the L1 required a greater effort and higher L2 fluency among her young participants when completing post-viewing tasks. Lwo and Lin (2012) compared L1 and L2 text, using a multimedia animated reading tool to explore the effects of different type of on-screen text (L1, L2, L1+L2 and



none) on vocabulary and reading comprehension with Grade 8 learners. They found that the effects of different modes on scores depended on learners' L2 proficiency, that for the lower-proficiency learners having L2 or L1+L2 subtitles was more beneficial, and that learners relied on visual information for comprehension. Naghizadeh and Darabi's (2015) study on L2 vocabulary with intermediate-level 15-17 year-olds reported that learners in the captions groups learnt significantly more word meanings than those on the subtitles group, who in turn had similar results to the non-subtitles group. Peters et al. (2016) conducted two experimental studies on the effects of L1 and L2 text on vocabulary gains for 17-18 year-olds (intermediate and low intermediate). Their results showed that, even if gains were low, captions had the potential to increase form learning, and that the captions group outperformed the subtitles group in form recall and meaning recognition and recall.

Altogether, the above studies suggest that, regardless of subtitling mode, length of exposure to the input, or proficiency, learners at these young ages benefit from exposure to audio-visual input enhanced with L1 and L2 on-screen text. As with older learners, results seem to indicate that captions are more adequate for older/more proficient young learners, while subtitles would be more appropriate for younger / less proficient children.

### **2.5.2. Frequency of occurrence**

Another factor that can affect word learning, independently of the language of the on-screen text, is the number of encounters with an item within the input. Research on vocabulary learning through reading has provided evidence that having repeated encounters with unknown words has a positive effect on vocabulary learning (e.g., Hosrt, Cobb & Meara, 1998; Pellicer-Sanchez & Schmitt, 2010), a finding also reported in studies on listening (e.g., Vidal, 2003, 2011; Van Zeeland & Schmitt, 2013) and more

recently in viewing studies (e.g., Peters et al., 2016; Peters & Webb, 2018; Rodgers, 2013). The more frequent words are learnt better than the less frequent words. Similar to the conundrum of establishing the lexical coverage needed for adequate comprehension, it is also difficult to establish the exact number of encounters needed to learn a word incidentally, and the number varies according to the input mode (written, aural) and the aspect of language learning being assessed (e.g., recognition vs. recall). Also, research has shown that this variable might interact with a number of other variables, such as learners' vocabulary size (e.g., Peters & Webb, 2018), and that it might be less relevant than other factors when learning through audio-visual input (Peters & Webb, 2018).

In reading research, the number of encounters necessary to learn an unknown word oscillate between 8 and 10, and learning of spelling requires fewer encounters than other types of word knowledge (e.g., Pellicer-Sanchez & Schmitt, 2010; Webb, 2007). Waring and Takaki (2003) found that, for example, for form recognition, learners needed to encounter a word at least eight times to have a 50% change to recognize it three months later, while for meaning recognition there was only 10% to 15% change to do so even if the word was met more than 18 times. If the number of encounters was below five, the chance was close to zero.

Pigada and Schmitt (2006) explored acquisition of form (spelling), meaning, and grammatical features. They found that spelling was acquired with few exposures, while meaning did not seem to be affected by frequency, except on the extremes – with 1 occurrence 3.4% of words were recalled, while with 20 occurrences the percentage rose to 60%. In grammar, there was a steady increase in learning as frequency increases. Pigada and Schmitt concluded that 20 encounters are needed to be able to learn all three knowledge dimensions. Pellicer-Sánchez and Schmitt (2010) also explored acquisition of form and meaning, and found a significant increase in learning after 10 occurrences,

with gains of 76% in form recognition, 63% in class learning, 84% in meaning recognition, and 55% in meaning recall. Contrary to Waring and Takaki's (2003) findings, however, they found that meaning recall required fewer repetitions than form recognition.

Brown, Takaki and Donkaewbua (2008) compared the effect of frequency of occurrence under three conditions: reading-only, reading-while-listening and listening-only. Brown et al. found that, for reading-only and reading-while-listening modes, with 7 to 9 encounters meaning recognition rates were around 45%, but for meaning recall dropped to 10% and 14%, respectively. In the listening-only mode, with 10 - 13 encounters there was a 36% recognition rate, which dropped to 1.5% for meaning recall, suggesting that picking up words from listening is considerably slower than from reading, and suggesting that much more than 20 encounters are needed – the authors argue that maybe 50, or up to 100 might be necessary. Brown et al. also conclude that for long-term retention, the number is much higher than 7 - 9 times, and closer to 30 - 50 times – even in reading. In contrast, Rott (1999) compared vocabulary gains through reading after 2, 4 and 6 encounters, and found that with 2 encounters there were already significant gains, and that 6 encounters – which led to higher gains than 2 and 4 – might be enough. Another study comparing listening and reading was Vidal's (2011) study, in which she found that gains were significantly higher in reading than in listening. Vidal also found, however, that the gains from listening were retained better. Frequency of occurrence significantly affected both modes, but the effect was smaller in the listening mode. That is, more repetitions were needed in listening (5 - 6) than in reading (2 - 3).

Van Zeeland and Schmitt (2013) investigated three dimensions of vocabulary learning (form, meaning and grammar) through listening, and compared the effect of different number of exposures (3, 7, 11 and 15). They found that frequency of occurrence did not have a strong effect on word learning, particularly in the case of word meaning acquisition. Larger gains were made in form and grammar than in meaning, but meaning

was less sensitive to attrition than the other two. While there was a significant advantage of 7 occurrences over 3 in form and grammar in the immediate post-tests, there was only a 47% chance the learner could recognize the word when heard again, and this effect disappeared with time. For meaning learning, there was no effect of repetition. In line with findings from Pigada and Schmitt (2006) and Brown et al. (2008), the authors suggest that more than 20 encounters might be needed. On the other hand, it has also been pointed out that other word-related factors might predict learnability better than frequency. For instance, in Vidal's (2003) study on academic listening she found a positive effect of repetition, but this factor predicted less than the other three included in the study (namely word form predictability, type of word and type of elaboration).

Studies on audio-visual input propose a lower number, suggesting that 5 occurrences might be enough (e.g., Webb & Rodgers, 2009) as the presence of visual support might compensate for the lower number of repetitions. Because of the nature of the audio-visual materials – which do not allow for modification of number of encounters –, studies on this mode of input have taken frequency of occurrence into account as an explanatory factor rather than aiming at establishing an exact number of necessary occurrences. The few viewing studies that have looked at this aspect, however, do suggest that frequency of occurrence has a positive effect on incidental word learning. Rodgers (2013) found a small but significant correlation between frequency of occurrence of target items and word learning gains in a demanding test of meaning recognition, although the correlation disappeared with a less demanding test format. Peters et al. (2016) also found that frequency was positively correlated with word learning. In the first experiment, they found that the effect was the same in both word form recognition (+10%) and meaning recall (+11%), whereas in the second experiment the effect of frequency was mediated by the interaction of frequency and learners' vocabulary size, with higher odds of learning a word when both parameters increased. Peters and Webb (2018) also found that

frequency was positively related with word learning. In experiment 1, they found that per each additional occurrence of the word in the input (i.e. a full-length documentary), the odds of recalling its meaning were 25% higher. In experiment 2, the odds of meaning recall were 20% higher when frequency of occurrence increased – per every five occurrences of the target item, the changes of recognising it doubled. Comparing results from both experiments, the effect of frequency was slightly stronger for meaning recall than meaning recognition.

A recent meta-analysis by Uchihara, Webb and Yanagisawa (2019) addresses the complex relationship between repetition and second language incidental vocabulary learning. Results showed that there was a medium effect of repetition ( $r = .34$ ) on incidental word learning, and that differences between studies were due to a number of variable such as learners' vocabulary size, testing format, and treatment variables (e.g., image support, spacing between encounters). One of the reasons for the absence of correlations between frequency of encounters and vocabulary gains might be the restricted range of encounters included in the studies (Pellicer-Sanchez, 2017). As summarized by Uchihara et al.'s (2018):

“The range in number of encounters with target words has varied from study to study: 1 to 3 (Hulstijn, Hollander, & Greidanus, 1996), 4 to 8 (Pellicer-Sánchez, 2017), 1 to 7 (Chen & Truscott, 2010), 2 to 17 (Horst et al., 1998), 5 to 54 (Rodgers, 2013), 1 to 70 (Webb & Chang, 2015), and 1 to 209 (Saragi et al., 1978). The absence of significant correlations between frequency of encounters and learning may be attributed to a restricted range in the number of encounters (Pellicer-Sánchez, 2017). Crucially, a limited range of encounters (e.g., 1 to 3) tend to underestimate the resulting correlation, since the data stem from a mere fragment of the full-scale “true” relationship that would be detected with a wider range of encounters (e.g., 1 to 209) (see Thorndike, 1949 for discussion of this issue in greater detail). Therefore, it is reasonable to hypothesize that the wider the range in number of encounters, the higher the correlation.” (Uchihara et al., 2019: 8)

Another moderator variable identified in Uchihara et al.'s (2019) meta-analysis was spacing between encounters, which refers to whether encounters with a lexical item are grouped together (i.e. massed) or spaced out over time (i.e. spaced). It was found that the effect of frequency was stronger in studies conducted in massed conditions, suggesting that repeated encounters with a short time span would be more beneficial for incidental word learning.

### **2.5.3. Explicit vocabulary instruction**

A way to optimize the effectiveness of vocabulary learning through TV programmes is to involve intentional or explicit learning, on the basis that incidental and intentional learning are complementary approaches that can be integrated (Schmitt, 2010). Research in the area of extensive reading suggests that learning rates can be increased by deliberately focusing attention on vocabulary (Elley, 1989; Hulstijn, 2013, Nation, 2015), and that pre-directing attention to word-forms (e.g., with input enhancement) yields higher gains in vocabulary (Lee, 2007). Research on listening with advanced organizers also showed that including some kind of pre-listening support had a positive effect on comprehension (e.g., Chang & Read, 2006; Chung, 2002), and that “simply playing videos as means of developing listening skills was not enough” (Chung, 1999: 301). Data from an eye-tracking research on incidental learning through reading showed that there is a relationship between the amount of time spent in a word and the ability to recognize it later, that is, that eye-fixations can predict vocabulary learning (Godfroid, Boers & Housen, 2013). If attention is pre-directed to certain words, learners might spend more time on them because words are made more salient, and if we assume that more attention leads to more learning (Boers, 2018; Robinson, Mackey, Gass & Schmidt, 2012), the potential to learn those words might increase.

However, guessing meaning from context – even with the additional help of images – can be challenging due to the real-time nature of this type of input (and particularly when viewing it with captions, with no access to L1 translations). Providing explicit access to the meaning of unknown words in the form of glossaries or pre-viewing activities may aid learners making an initial form-meaning connection (Chai & Erlam, 2008; Montero-Pérez, Peters & Desmet, 2018; Sydorenko, 2010; Webb, 2010c; Yang, 2014). Webb (2010b; see also Webb & Rodgers, 2009a) investigated the potential of pre-teaching low-frequency words to increase comprehension by analysing the lexical profiles of several TV series. Webb pointed out that television programmes may be too demanding for lower level learners because they do not have the vocabulary necessary to understand the content, and he suggested pre-learning unknown topic-related words in a specific television programmes to improve comprehension and vocabulary learning, but to date no study has looked at the effects of teaching (or just directing the attention to) target expressions as part of a classroom intervention.

To the author's knowledge, no studies have empirically investigated the effects that having access to (form and) meaning may have on the learning of lexical items through audio-visual input. An exception is Montero-Perez, Peters and Desmet's (2018) study, which investigated vocabulary learning through two enhancement techniques (i.e. test announcement (yes or no) and type of captioning (full-captioning, key-word captioning, glossed key-word captioning, no captioning)). This last type of captioning provided access to meaning while watching the audio-visual material. Results showed that, while the test announcement did not yield differences between groups, type of captioning did, with the group with access to glosses significantly outperforming the other three in meaning recall. The glossed captions group recalled around 7 word meanings (out of 18), while the other only recalled 2-3 word meanings. The authors conclude that "if we want to stimulate the intentional learning of new words, form-focused activities before or

after viewing videos could be a more effective approach” (Montero et al., 2018: 19). In fact, there are two studies that have included such pre-viewing activities (Bravo, 2008; Gesa, 2019), but, since their aim was not the comparison between teaching and no teaching, no conclusions can be drawn about the specific effects of instruction in those studies.

Bravo (2008) investigated vocabulary learning and content comprehension through ten 15-minute episodes of an American TV comedy. Participants were 77 Grade 9 students (aged 13-14, with A2/B1 proficiency level) divided into two experimental groups: one viewed the episodes with L1 subtitles while the other viewed them with L2 captions. For each episode, participants were given a 20-multiple-choice-item task, including 10 items on content comprehension and 10 items on lexical phrases. The task was handed out prior to the viewing, and participants were asked to indicate which lexical phrases they were familiar with. The lexical phrases were highlighted to attract attention to them. Besides the post-viewing tasks, four vocabulary consolidation tests (CT) were administered. CT1 and CT2 were given after episodes 5 and 10 respectively, and included 25 lexical items (5 from each episode). CT3 was given three weeks after the end of the intervention, including 50 lexical items, while CT4 was given three months after and included the same 50 items.

Because results from the post-viewing task were presented together, it is not possible to distinguish results between comprehension and vocabulary. Overall, Bravo reported a marked improvement after episode 4 for both language groups. The captions group performed slightly better from episode 1 to 3, while the subtitles group outperformed its counterpart from episode 4 onwards, with significantly better results in two episodes. As per the consolidation tests – testing lexical items – there was an improvement from CT1 to CT2, suggesting a training effect. The percentage of correct responses was very similar for both experimental groups, with captions (82.1%) performing better than



subtitles (78.3%) in CT1, while differences levelled off in CT2 (80.8% for subtitles and 80.5% for captions). As for the first delayed post-test (CT3), the subtitles group (68.2%) outperformed again the captions group (66.1%), but differences levelled out by CT4, with both captions (65.7%) and subtitles (65.2%) scoring similarly. Bravo concludes that captions were more challenging because learners had to process the utterances in the L2 and then reformulate their meaning in the L1, which takes longer time. However, “the difference [between the groups] was marginal and these preliminary results should not be interpreted as a conclusive answer concerning the most effective viewing mode for foreign-language learning or improvement” (Bravo, 2008: 166).

The second study including pre-viewing activities was Gesa (2019), who carried out a longitudinal classroom-based intervention watching several episodes of TV series with Grade 6 (A1), Grade 10 (A2) and university students (B1). Primary school students and secondary school students completed a 10-month intervention, viewing 22 and 24 full-length episodes (respectively) of a 20-minute TV series, while university students did a 3-month intervention (viewing eight 20-minute episodes of an American TV comedy). In each age group, there were two conditions (i.e. experimental and control). Experimental groups watched the episodes with subtitles in the L1 (Grade 6) or captions in the L2 (Grade 10 and university), and completed a short pre-viewing activity, a while-watching activity and two post-viewing activities – namely a vocabulary task and a comprehension test. The pre-viewing activity was aimed at teaching five target lexical items appearing in the episode, using a combination of tasks such as word-searches or fill-in-the-blanks. Control groups completed pre- and post-viewing tasks but without watching the episodes (instead, they had regular EFL classes).

Learning of the target items was assessed by means of a pre- and post-test administered at the beginning and end of the each term. For secondary school participants, results showed that the experimental groups outperformed the control

groups in both form (Grade 6: 18.4% vs 14.8%; Grade 10: 35.2% vs 27.8%) and meaning recall (Grade 6: 10.5% vs 9.6%; Grade 10: 20.1% vs 12.6%), but differences did not reach statistical significance. Results also showed that Grade 10 outperformed Grade 6, which is not surprising considering the proficiency and age gap between the two grades. For university students, no significant difference was found between the experimental and control groups, suggesting that it was the task – and not the video - that significantly contributed to word learning.

One study compares grammar learning with instruction and without instruction through a movie with reversed subtitles. Van Lommel, Laenen and d'Ydewalle (2006) carried out two experiments with 11-year-olds and 17-year-olds in which they tested five Esperanto grammatical rules through a 40-item multiple-choice test (8 per rule). Half of the items appeared in the movie, whereas half were new to the students. In the first experiment, each age group was assigned to one of the four experimental conditions depending on the variable *movie* (with vs. without) and *advanced rule presentation* (with vs. without). Results showed that older participants and participants who received instruction performed significantly better. Presentation of the rules beforehand enhanced performance especially for older students compared to the younger ones, indicating an interaction between age and advanced rule presentation. Without instruction, however, the presence or absence of the movie did not affect performance in either age group, suggesting that watching the movie itself does not produce acquisition of grammar, which would fall in line with findings by Gesa (2019). Participants who watched the movie, however, performed better in those items that appeared in it, suggesting that the movie had an effect. The second experiment followed a similar procedure, but also included differences in the instructions (movie with test announcement, movie without test announcement, and no movie). Overall, results from both experiments indicated that grammatical rules were not acquired through the movie alone, that there was a strong

effect of pre-teaching and that this effect was stronger in older children. Although the effect of the movie was not statically significant, participants who watched it performed slightly better, and items that had appeared in the movie received higher scores. Teaching the rules in advance was equally beneficial for all items.

In sum, while there has been an increase in studies focusing on vocabulary learning through audio-visual input, there is a scarcity of research on explicit vocabulary teaching to boost vocabulary acquisition through extensive viewing. Neither has research examined whether attention on other aspects of language learning might come at a cost for comprehension, which can be a common activity within the FL classroom setting. If learners are required to focus their attention on, for example, vocabulary, this might come at a cost for comprehension, as attentional resources are limited. While research on the use of advanced organizers has shown that providing pre-listening activities has a generally positive effect on comprehension (e.g., Elkhafaifi, 2005) – as they seem to help listeners activate their prior knowledge (top-down processing) (Vandergrift, 2007) – explicit focus on vocabulary yields conflicting results. Chang and Read (2006), who investigated various forms of support for listening comprehension, found that vocabulary instruction was the least effective one – regardless of proficiency level. In another study Lee (2007) explored the effects of textual enhancement and found that, while vocabulary improved, overall comprehension decreased. This can be seen as a special case of cognitive overload of the verbal channel (see the Cognitive Load Theory in multimedia learning; Mayer & Moreno, 2003; Sweller, 1999, 2005) which occurs when learners' cognitive processing exceeds the available cognitive capacity.

#### **2.5.4. Vocabulary retention**

Several studies on incidental vocabulary learning have assessed the extent to which the vocabulary learnt through an exposure to audio-visual input can be retained afterwards, revealing differences between retention rates depending on the type of on-screen text participants had access to. Baltova (1999) tested retention of lexical items two weeks after viewing a brief documentary with captions, reversed subtitles and without subtitles. She found that participants who had viewed the video with captions had learnt and retained a significantly higher amount of vocabulary than the other two conditions. In Letorla's (2012) study, investigating learning through a subtitling task, it was found that learners with access to subtitles had higher scores than the control group (without subtitles) when tested a week later. Nagira (2011) compared vocabulary retention rates between two captioned and non-captioned short clips, also a week later, and found that there was low decay in both condition. Vulchanova et al. (2015), on the other hand, did not find long-term effect when testing participants four weeks later. Peters and Webb (2018) assessed vocabulary retention a week after administering the post-test, and found that scores were higher in the delayed-post-test than in the immediate post-test, and that there were a relatively large number of item learnt between the two testing times. The authors argue that, while it was positive that the test drew learners' attention to new words, from a research perspective gains in the delayed-post-test could not be only attributable to the treatment.

In spite of the methodological differences between the studies, findings suggests that a proportion of the vocabulary acquired during the viewing can be retained afterwards. Studies have generally investigated, however, short-term retention, with delayed-post-test being administered between a week and a month after the post-test. The long-term effects of subtitles and captions on vocabulary retention are, however, still unclear (Vandergrift, 2011).

## **2.6. Attitudes towards OV input**

When assessing the extent to which language learners can benefit from exposure to OV (Original Version) TV input<sup>5</sup> with subtitles or captions, one needs to bear in mind – besides the age and proficiency level of the learners – other factors that might affect input processing, such as their attitudes towards this kind of input or how familiar they are with it. This seems especially relevant in a context such as Spain, where – as previously mentioned – learners are not usually accustomed to viewing L2 television.

### **2.6.1. Familiarity with viewing captioned and/or subtitled OV input**

Familiarity with viewing OV audio-visual material may have an impact on the viewing process itself. A European survey (2011) carried out in 33 countries and with 11,000 respondents found that younger people (aged 12-25) preferred subtitling over dubbing, but with a significant exception in dubbing countries (such as Spain), where even young citizens preferred dubbing to subtitles – primarily out of habit. Respondents from subtitling countries were more adept at quickly developing strategies to take advantage of them compared to those coming from dubbing countries (Vanderplank, 1988). This suggests that familiarity with the use of on-screen text (either in the L1 or the L2) may play a role, and that learners who are used to them might be able to benefit more.

Koolstra and Beentjes (1999) found a positive relationship between vocabulary learning and the frequency with which children watched subtitled TV programmes. However, Taylor (2005) found that, with captions, beginner students with little background in reading and listening in the foreign language found it difficult to attend to

---

<sup>5</sup> OV refers to input in its original version in spoken English (not dubbed).

the three channels and were confused or distracted by the presence of captions. Taylor emphasized, though, that what was surprising was not only that beginner learners could not simultaneously process the three channels, but that learners who had only two more years of language study were capable of doing so. Pujadas & Muñoz (2017) interviewed a group of secondary school students from the same sample as those in the current study who had been watching TV series in the classroom for six months, and found that learners reported a change in their viewing habits at home – moving from dubbed to subtitled TV watching, and from subtitles to captions. Students also reported that by the end of the year they understood the series better as they got used to actors and their voices, a finding reported in previous studies too (e.g., Rodgers, 2013). As suggested by Danan (2004), captions may lead to significant improvement in comprehension when learners are trained on relevant strategies, and additional practise may lead to active knowledge of the recently encountered words (Danan, 1992).

Training effects have been hypothesised in different studies. For instance, in their study with young children, Koolstra, Van der Voort and Van der Kamp (1997) found that watching captioned/subtitled television programmes facilitated – over time – the development of decoding skills through the reading practise resulting from reading subtitles displayed on television (the on-screen reading hypothesis). Because the words are exposed only for a short moment and then replaced by new on-screen text, children are constantly trying to keep up with the presentation tempo of the subtitles. Hasan (2000) found learners who were used to their non-native teacher accent said to find it difficult to understand other speakers' accents, but the author argues that getting used to other accents might only be a temporary problem, which might be solved through practise.

Frequency of watching OV input can provide an estimate of how familiar a learner is with this type of input. From data collected through questionnaires, Kuppens (2010)

reported that Grade 6 students who frequently watched subtitled English programmes (before formal instruction) performed significantly better in English tests. The influence of frequent watching of audio-visual material over other types of out-of-school exposure has been observed in several studies as well. Lindgren and Muñoz (2013) found that watching movies had the strongest explanatory power on the listening and reading comprehension skills of a very large group of 10-11 year-old learners in seven European countries. Similarly, in a study comparing Danish and Spanish learners of English (ages 7 and 9), Muñoz, Cadierno and Casas (2018) found that Danish children's more frequent viewing of OV audio-visual material contributed to compensate for the effects of their comparatively lower amount of formal instruction to a greater extent than other activities such as gaming or listening to music. Peters (2018) showed that 40% of surveyed students (15-16 year-olds) watched TV several times a week, compared to the 1% who indicated reading books with the same frequency.

In a recent study, Vanderplank (2019) investigated whether greater or lesser use of captions affected learners' performance on a language test, but caption use did not predict it. Instead, "prior experience with using films to aid language learning or familiarity with watching with captions were the most important factors" (Vanderplank, 2019: 10).

### **2.6.2. Attention and enjoyment**

Other factors that might play a role in understanding TV input are attention to and enjoyment from the TV series themselves. A concern might simply be whether learners are paying attention to the input or not, especially when research is classroom-based and TV viewing might be seen as a leisure-oriented activity (Vanderplank, 2016b). In contrast, results from a survey about attitudes towards TV input in the L2 classroom indicated that – independently of age and language skills – learners found TV viewing

more enjoyable and engaging than traditional listening activities (Pujadas & Muñoz, 2017).

Motivation and interest in the content can also be important factors in understanding the spoken discourse. Students can often find it difficult to understand a spoken text when they are not interested in the topic (Hasan, 2000). In a study with Grade 11 participants, Baltova (1999) found that students' attitudes towards scientific documentaries were significantly (and positively) correlated with students' performance on the language tests, independently of the experimental condition in which participants had watched the videos (the study included different types of subtitles).

### **2.6.3. Attitudes towards L2 television and on-screen text**

Learners' perceptions and attitudes towards captions and subtitles can affect how language learners make use of them (e.g., Pujolà, 2002). An increasing number of studies on language learning through audio-visual input investigate – besides the linguistic benefits of subtitles and captions as measured by comprehension and vocabulary tests – how learners perceive the usefulness of on-screen text for language learning purposes. Through questionnaires and interviews, studies have gathered valuable qualitative information on the usefulness of and preferences for either captions or subtitles. This has allowed for a triangulation of results and the addition of a qualitative perspective to the quantitative findings, and it has provided possible explanations for differences between learners.

Overall, studies report positive attitudes towards using audio-visual materials (usually with captions) in the language learning context. Kim (2015) inquired about participants' attitudes in a study with video only, and found that learners considered video input as an



interesting, motivating and effective language learning tool – independently of their proficiency level –, as well as being useful for learning non-verbal expressions and grasping cultural aspects. Studies including captions have also found positive reactions to this media, either with or without text support (e.g., Chai & Erlam, 2008; Kim, 2015; Rodgers, 2013; Vanderplank, 1988). Students have reported that captions can help reduce anxiety when faced with input that might be beyond their perceived language skills, and lead to greater self-efficacy and confidence as learners understand the video better (Guillory, 1988; Vanderplank, 1988). Attitudes towards captions vary, however, depending on the type of on-screen text (e.g., Guillory, 1998; Montero-Perez, Peters & Desmet, 2014) and learners' L2 language skills (e.g., Park, 2004; Taylor, 2005).

#### *2.6.3.1. Simultaneous processing of image and on-screen text*

The effect that captions have on learners' attention to the input can be viewed as both beneficial and distracting (or attention depleting). Captions are sometimes viewed as a "crutch" that reinforces and confirms what is taken in aurally, but, at the same time, might sometimes lead to only read and not listen (Winke et al, 2010).

Studies on eye-tracking report that viewers tend to read the on-screen text, regardless of their knowledge of the language in which they are displayed (d'Ydewalle & de Bruycker, 2007; d'Ydewalle, et al., 1991), their familiarity with on-screen text (d'Ydewalle et al., 1991), or the availability of the soundtrack (d'Ydewalle, van Rensbergen, & Pollet, 1987; Van Lommel, Laenen, & d'Ydewalle, 2006). Students themselves also usually reported to prioritize the reading of captions when watching audio-visual input, while some report to be unable to pay attention to both sound and pictures at the same time (Chai & Erlam, 2008). In Stewart and Pertusa (2004), 65% of participants claimed to pay more attention to the audio than the captions. In Taylor's (2005) study, 74% of

participants said they attempted to listen to the audio, but around 25% recognized they concentrated on the captions and ignored the sound. Sydorenko (2010) reported that participants paid the most attention to the captions, then the video, and finally the audio. At the same time, however, students indicated that the video was more helpful than the captions, and that they tried to match visual images with words. As the author suggests, the reason might be that captions require a minimum L2 competence, whereas learners have no problem in processing visual information. Sydorenko also noted that it was especially beginning learners who considered video images more useful than captions, maybe due to the fact that they could only understand parts of the dialogue. Thus, attention to one channel or another can be mediated by the learners' proficiency level. In Taylor's (2005) study, the percentage of participants who reported the ability to use all three channels (audio, text and sound) was of 50% for more advanced learners, but only 23% for beginner learners.

#### *2.6.3.2. Perceived usefulness*

Besides their ability to process simultaneously all three channels, students have a wide range of perceptions about what can they learn while viewing audio-visual media, and about the usefulness of on-screen text. Captions aid with the decomposition and analysis of language (Winke et al, 2010), and facilitate "tuning in" to fast, authentic speech and unfamiliar accents (Vanderplank, 1988). In Vanderplank's (2019) study, more than half of participants recognized that it would be very difficult or even impossible to fully understand the input without captions. Captions are found useful for recognizing and learning new words (e.g., Montero-Perez et al, 2013) and help match the aural and written forms (e.g., Chung, 1999, Stewart & Pertusa, 2004). Captions are also perceived as useful for spelling, retention and recall (Vanderplank, 1988). Common guessing

strategies reported by learners are: recognizing words that are similar to L1, using the roots of known words, paying attention to verbal context, paying attention to grammar (Sydorenko, 2010).

Reports from different studies suggest that proficiency also plays a mediating role in learners' perceptions. Pujolà (2002) found that more advanced learners tended to see captions "as a backup to their listening activity" (Pujolà, 2002: 254), whereas lower-level learners considered captions fundamental for comprehension. It seems, however, that beginner learners are the ones with more difficulties to make use of captions. In Vanderplank (1988), learners below an intermediate level or slow readers found it difficult to use them. This was also found in Park's (2004) study, where the high-intermediate learners reported positive attitudes about key-word captioning, while intermediate and low-intermediate students found them distracting. Taylor (2005) compared strategy use of captions between first- and third-year college students, and found that experience with the foreign language made it easier to take advantage of them. Taylor reports that first-year students found captions more distracting or confusing, and that a larger proportion of third-year students reported the ability to use all three channels.

Fewer studies have looked into L1 subtitles specifically. In a study on multimedia annotations including visual and verbal annotations, Jones (2003) found that while L1 annotations were helpful, they did not encourage deep thinking. Stewart and Pertusa (2004) reported that more than 50% of the participants found it difficult to learn new words through L1 subtitles because they could not see the words written, and half of the participants also indicated that they would prefer having captions to subtitles. The authors conclude that "concerns that students might be lost without the aid of English subtitles appear to be unfounded. Many of these students clearly recognized benefits for their language learning that are worth the extra effort required to view films with Spanish subtitles [L2]" (Stewart & Pertusa, 2004: 440).

Chung (1996, 1999) explored the perceptions on video-aided English teaching with captions combined with different type of advanced organisers. In both studies, it was found that most participants (77% and 90% respectively) said that advanced organizers helped them understand, even when they did not understand certain vocabulary items. Participants did not consider that pre-viewing activities were a waste of time, although knowing the plot in advance seemed to lessen curiosity. Most students (83%) also felt that the combination of both captions and advanced organizers enhanced their listening comprehension (Chung, 1999), but that preferences depended on proficiency: more advanced students might not need both, but lower level students might. Advanced organizers were found especially useful for providing explanation of vocabulary, idioms and phrases prior to viewing (Chung, 1996).

#### *2.6.3.3. Changes in attitudes over time*

The majority of the aforementioned studies report attitudes towards audio-visual input and on-screen text at a given point in time, and data are collected either before or after the experiment itself. Few studies have looked at how those attitudes may change over time, after (relative) extensive exposure to audio-visual input (Vanderplank, 2019). One such study is Rodgers's (2013), who looked at learners' perceptions during and after an extended exposure to TV series. Students reported that learning through audio-visual input had a useful effect on their English ability in general and that their listening ability improved. They also found it to be a good use of class time and enjoyable experience. Rodgers also assessed students' enjoyment after each of the 13 episodes viewed, and found that the enjoyability of, benefits from and usefulness of language learning through TV viewing significantly increased from the first to the last episode (drastically after the first episode, and levelling off as more episodes were viewed).

In a recent study, Vanderplank (2019) investigated – through the use of personal diaries – changes in motivation, strategies and attitudes by 36 FL learners after viewing captioned and non-captioned movies for an extended period (6 or 12 weeks). Participants had choice and control of the viewing material, which included optional captions. Data from the diaries revealed that 30% displayed a positive attitude from the start, while the 11% who expressed negative comments at the beginning changed them over time. Data also revealed that the more movies they watched, the less they noticed the on-screen text, while their confidence to watch movies without captions increased. Participants reported that changes even occurred within the same film (e.g., “intriguing film, I stopped paying attention to whether or not I was reading the captions from about half way through” (Vanderplank, 2019: 7)), and many could become accustomed to specific people’s speech during the course of the movie.

#### 2.6.3.4. *Younger learners*

Few studies have explored learners’ attitudes and perceptions with a non-university population, but it has generally been found that EFL learners are also positive about the use of L2 television.

However, to the author’s knowledge, only one study including an extended period of exposure to audio-visual input has looked at the perceptions of younger learners. Bravo (2008) asked Grade 8 participants to complete a questionnaire about their opinions on the intervention (i.e. viewing 10 episodes of a TV series with captions or subtitles) – although only participants in the captions conditions were asked to do so. Students reported that captions help them improve their written and spoken English, aid spelling and pronunciation, and that they are useful for learning vocabulary and expressions, as well as how sentences are constructed. Captions also helped them compensate for

missing pieces of information and improved concentration at watching. Interestingly, one student acknowledged that “being forced to read English” made him or her learn more (Bravo, 2008: 164). On the other hand, students also enumerated some disadvantages, such as not having time to read the captions – due to the fast pace and the fact that they are in the foreign language. They also reported that at the beginning “this technique it’s puzzling and confusing” (Bravo, 2008: 164), suggesting that they could cope better with captions after some time.



## Chapter 3. Rationale

---

### 3.1. Introduction

Research evidence indicates that TV series are a potential and suitable tool for language acquisition, as they expose learners to large amounts of comprehensible input and provide opportunities for incidental vocabulary learning. Previous research has suggested that TV series can be enhanced by adding on-screen text, either in the form of captions or subtitles, making the input more accessible for language learners while providing numerous benefits for language learning. The advantages of using either language seem to depend, however, on which aspect of the language is the focus on (e.g., content comprehension, vocabulary learning), as well as the learners' proficiency and age. The majority of studies, nonetheless, have been carried out with an adult population and at an intermediate L2 level. Also, studies have generally focused on captions – which have been shown to be more demanding for learners with limited L2 competence.

Studies on audio-visual input have generally directed their attention to incidental learning of vocabulary. Yet, it is well established that having explicit instruction can lead to greater language gains, and therefore providing explicit vocabulary instruction prior to viewing could contribute to higher learning rates. Nevertheless, studies investigating the role of pre-teaching in the context of audio-visual input are still scarce. It has also been argued that the benefits yielded by audio-visual materials for comprehension and vocabulary learning can increase if learners are exposed to it in large quantities (i.e. by watching consecutive episodes of a TV programme). A large proportion of studies, however, have used educational videos and short clips, which are not fully representative



of what people would normally choose to watch. There are also few studies including more than one video or longitudinal studies including several videos across more than one session.

In addition, research has also suggested that there are other factors – besides the type of on-screen text and the addition of instruction – that may mediate learning in this context, such as frequency of encounters with the words and imagery. Prior studies have found that words encountered more often are better learnt, and that words tend to re-occur within the same TV programmes. It has also been suggested that words that have image support are more likely to be learnt. Finally, learners' attitudes towards this type of input may also play a role on the effectiveness of L2 television for language learning.

### **3.2. Research questions**

The present doctoral dissertation will try to contribute to filling in these gaps with data from a one-year-long classroom intervention with adolescent language learners, who watched 24 episodes of a TV series under four experimental conditions depending on the language of the on-screen text (captions or subtitles) and pre-viewing instruction (with or without). The dissertation is divided into four main areas of interest: viewing comprehension, vocabulary learning, the role of frequency and imagery, and learners' attitudes.

The corresponding studies were designed to answer the following research questions:

### **Study 1. Viewing comprehension**

1. To what extent does the language of the on-screen text affect comprehension of TV series? To what extent is comprehension also affected by:
  - a. Instruction-related factors (i.e. explicit focus on vocabulary items)?
  - b. Learner-related factors (i.e. general proficiency, vocabulary size, familiarity with OV, attention to and enjoyment from the series)?
  - c. Test-related factors (i.e. item format, type and difficulty)?
  - d. Episode-related factors (i.e. lexical coverage)?
2. Does comprehension of the episodes improve after an extended exposure to TV series (i.e. viewing 24 episodes over a period of eight months)?

### **Study 2. Vocabulary learning**

1. To what extent can L2 vocabulary (form and meaning) be learnt through an extended exposure to TV series in the classroom?
2. To what extent is vocabulary learning through TV series affected by:
  - a. Explicit focus on vocabulary (i.e. pre-teaching target lexical items or not)?
  - b. Language of the on-screen text (i.e. captions or subtitles)?
  - c. Learners' proficiency level?
3. Does the percentage of vocabulary learning increase as more episodes are viewed?
4. Is there a relationship between comprehension (Study 1) and vocabulary gains?
5. To what extent is the L2 vocabulary learnt through an extended exposure to TV series retained in the long term (i.e. eight months after the intervention)?

### **Study 3. The role of frequency and imagery**

1. To what extent does frequency of encounters predict word-form and word-meaning learning in TV viewing?
2. Is the effect of frequency mediated by intervention-related variables (i.e. type of instruction, language of the on-screen text) and learners' proficiency?
3. To what extent does the imagery from videos support word-form and word-meaning learning?
4. Do explicit instruction and learners' proficiency mediate the effect of imagery on word learning?

### **Study 4. The learners' perspective**

1. To what extent do learners' preferences towards the use of on-screen text change after an extensive exposure to TV series in the classroom? Are there changes in the language of preference (i.e. subtitles or captions)?
2. What are the main reasons for viewing OV input with subtitles, captions or without on-screen text?
3. To what extent do viewing habits change after an extensive exposure to TV series in the classroom? Are there long-term effects?
4. What is the perceived usefulness of captions and subtitles as a language learning tool *before* and *after* an extended exposure to TV series in the classroom?
5. What is the feeling of learning from exposure to TV series in the classroom? Is it related to the language condition (i.e. subtitles or captions) in which participants watched the TV series?

## Chapter 4. Methodology

---

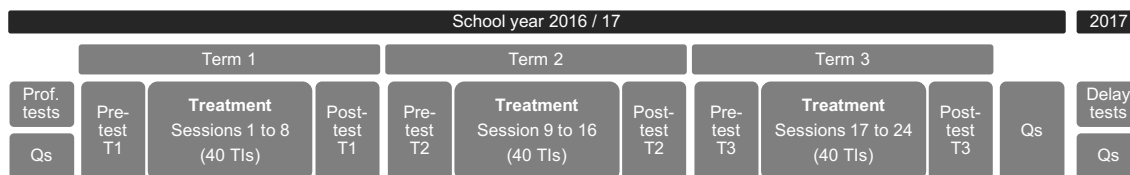
### 4.1. Introduction

This chapter is concerned with information about the intervention design, the participants, the TV series selected as viewing material, and the testing instruments used throughout the four studies of this dissertation – including standardized tests and tasks designed in-house. Information on specific aspects of the instruments will be further expanded within each corresponding study.

The intervention was carried out in a secondary school in the area of Barcelona. Four intact classes of Grade 8 students took part in the intervention watching the 24 episodes of a TV series (distributed in three terms) over the course of one academic year, with each class randomly assigned to a different experimental condition. All students were exposed to the same videos, but the treatment varied according to the language of the on-screen text and the type of instruction they received prior to the viewings. Content comprehension and vocabulary learning were assessed by means of a battery of tests. Data on individual differences was gathered through proficiency and aptitude tests. A set of questionnaires was administered before and after the intervention to collect background information and students' perceptions towards captions and subtitled audio-visual input. Additionally, individual interviews were also carried out with a subsample of participants. Figure 4.1 presents a general overview of the intervention.

A copy of all intervention-related materials – including vocabulary tests, pre- and post-viewing tasks, comprehension tests and questionnaires – can be found in Appendix B (in the attached CD).

Figure 4.1. Overview of the intervention



Prof. tests = proficiency tests; Qs = questionnaires; Delay tests = Delayed post-tests

## 4.2. Participants

### 4.2.1. Background context

In Spain, students start learning English as a compulsory subject in elementary school, at age 6 (Grade 1), with some schools starting as early as the pre-school stage (age 3-5). At the beginning of Grade 8, students attending state schools in Catalonia have typically received around 1,000 hours of formal English instruction. Students might have also had contact with English in Content and Language Integrated Learning (CLIL), depending on the school they attended to, in different amounts and for different subjects<sup>6</sup>. Despite the slow shift towards a more communicative approach, teaching English as a foreign language is still largely based on explicit grammar instruction (Gilbert & Muñoz, 2010), and outcomes are not fully satisfactory. In spite of the large amount of hours dedicated to the formal teaching of English at schools, Spain finds itself amongst the European countries with the lowest proficiency level in English, being the second to last in listening comprehension (European Commission, 2012). Because of the strong dubbing tradition in the country, foreign television audio-visual programmes are

<sup>6</sup> In Spain, the education system is guided by the Organic Law Education 2/2006 (Ley Orgánica de Educación LOE 2006). While the Organic Act of Education creates a framework for the educational system of the country at a national level, the autonomous communities regulate and adapt this law to their respective areas, and thus there are various EFL/CLIL models throughout the country (Caraker, 2016).

commonly dubbed into Catalan (in Catalonia) or Spanish, which means that – compared to other European countries with wide exposure to English through newspapers, TV and other types of audio-visual media – exposure to English in Spain tends to be limited to formal settings, unless learners willingly seek for other sources of authentic input.

#### **4.2.2. Sample of participants**

The original pool of participants for the present study were 106 secondary school learners in Grade 8 (65 female, 41 male) from a state school in the area of Barcelona. They were 13-14 years old (mean age 13.35) at the time the intervention started, and they were Catalan-Spanish speakers – most of them balanced bilinguals for whom both languages may be considered first languages<sup>7</sup>. The school was located in the outskirts of Barcelona, in a city of 82,902 inhabitants and with a mid-low socio-economic level (IDESCAT, 2018, URL: <https://www.idescat.cat/>).

Prior to the intervention, participants had a beginner proficiency level in English (A1), ranging from Pre-A to B1 according to the Common European Framework of Reference (as measured by the Oxford Proficiency Test), and a mean vocabulary size of around 2,000 words (as measured by the X-Lex test). The mean proficiency level of the participants in the sample is equivalent to the English proficiency level that students are expected to achieve at this age in Catalonia<sup>8</sup>. Data from the background questionnaires

---

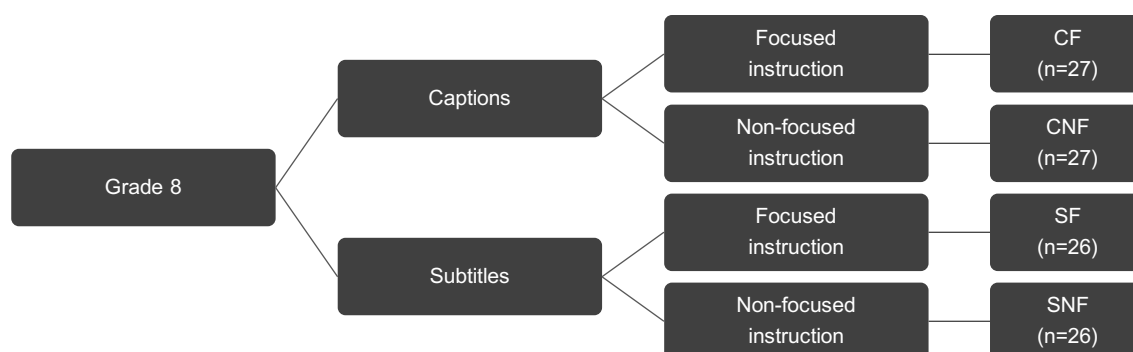
<sup>7</sup> In Catalonia, Catalan is the co-official language in conjunction with Spanish. Catalan is the language of instruction at schools, although Spanish is the majority language and its presence in the media is strong. People can have Catalan, Spanish or both as family language(s).

<sup>8</sup> According to Department of Education of the Generalitat de Catalunya, students should finish Grade 10 with a B1 level (Generalitat de Catalunya, 2015). Consequently, students should finish Grade 7 with an A1/A2 level, Grade 8 with A2 level, and Grade 9 with A2/B1 level. The sample of participants in this study can be considered, therefore, fairly representative of the larger population.

administered prior to the intervention revealed that around 55% of participants watched movies or TV series in English with L1 subtitles on a weekly basis and around 15% with L2 subtitles or no subtitles. More than 50% reported they found subtitles to be useful or very useful and only 4% considered them useless or annoying. Students' preferences and their viewing habits will be the focus of Study 4 (see Chapter 8).

In Catalonia, Grade 8 students have four lessons per week of English as a Foreign Language (EFL). The intervention took place during one of these weekly regular EFL lessons. Participants had already been randomly distributed in four classes by the school, and each intact group was assigned to a different experimental condition in order to explore the effects of the language of the on-screen text and the type of instruction on vocabulary learning and content comprehension. Two classes were assigned to the captions (English) condition, and the other two were assigned to the subtitles (Spanish) condition. One class in each language condition was also taught a set of target items (TIs) before viewing each episode (see below). Figure 4.2 shows the initial distribution of the 106 students into the four experimental conditions: captions with focused instruction (CF), captions with non-focused instruction (CNF), subtitles with focused instruction (SF) and subtitles with non-focused instruction (SNF).

Figure 4.2. Experimental groups and original sample numbers



### 4.2.3. Final sample of participants

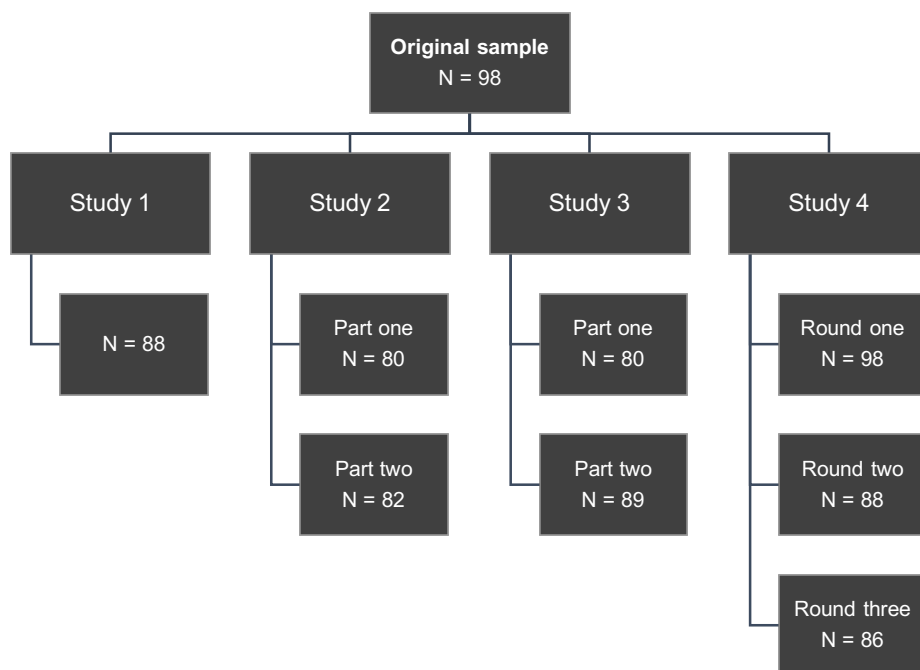
Because the study was conducted as part of the school's regular EFL lessons, all students enrolled in Grade 8 that year took part in the intervention<sup>9</sup>. However, from the original pool of participants, six students dropped the classes during the first term for various reasons (three were moved to a remedial class, two were expelled and one transferred to a different school), and two had an extremely low attendance record. From the remaining 98 students, only those who regularly followed the English course throughout the academic year and had a minimum of 85% attendance were included in Studies 1, 2 and 3, with a final valid sample between 80 and 98 participants depending on the study. For some of the analyses, participants who did not complete the language proficiency tests or the vocabulary tests had to be excluded as well. On the other hand, studies where data from a single term was analysed, attendance in that specific term was taken into account. The number of participants included in each study will be specified in the corresponding section of each study (see sections 5.3 [Study 1]; 6.3.1 and 6.4.1 [Study 2]; 7.3.1 and 7.4.1 [Study 3]; and 8.3 [Study 4]). Figure 4.3 summarizes the number of participants included in each study.

---

<sup>9</sup> Participants' parents were informed about the study through the school and students agreed that the data collected could be used for research purposes through passive consent.



Figure 4.3. Total number of participants in the four studies



#### 4.2.4. Terms

The terms *focused* and *non-focused* are used to refer to the condition in which learners are pre-taught the target items and are not, respectively, and these terms are preferred over the terms *intentional* and *incidental*. As discussed in prior sections, incidental learning can be defined from different perspectives, from not having forewarning of an upcoming post-test (Dörnyei, 2009; Hulstijn, 2003, Montero-Perez et al., 2018), to learning without intention to learn (Ortega, 2014), or learning as a by-product of another activity. Incidental learning is often defined by contrast to intentional learning, which occurs when “learners know that they will be posttested, or where their activities are explicitly focused on language learning” (de Vot et al., 2018: 3). The distinction with intentional learning is mainly centred, therefore, on the role of learners’ intention (Burton, García López & Esquileche Mesa, 2011). However, within the classroom context ‘it is difficult [...] to ensure that learners do not become intentionally

focused on learning vocabulary' (Malone, 2018: 3), independently of whether the test is announced or the activity revolves around other aspects of the language. Also, it is impossible to know what participants actually do (Hummel, 2010).

In the present study, learners who received no prior instruction on lexical items prior to viewing the episodes could be – to a certain extent – learning vocabulary incidentally. However, due to the longitudinal nature of the intervention, all learners were aware of the upcoming tests comprehension and vocabulary-focused tests after a couple of sessions, independently of the condition. Since it could be not affirmed that learners exposed to lexical items prior to viewing could have no intention to learn, the term “intentional” was not deemed appropriate either.

### **4.3. Audio-visual materials**

#### **4.3.1. The TV series**

The TV series selected for the study was *Fresh off the Boat* (Khan, Kasdan, Mar, Blomquist, Huang & McEwen, 2015), a sitcom produced by 20<sup>th</sup> Century Fox Television. Season 1 was originally released by ABC network in the United States, from February 4 to April 21, 2015, and contained 13 episodes. Season 2, with 24 episodes, was aired from September 22, 2015 to May 24, 2016. The series, based loosely on the life of American-Taiwanese chef Eddie Huang and his autobiography (*Fresh off the Boat: A Memoir*, 2013), follows the life of young Eddie and his family (see Figure 4.4) as they move from Chinatown in Washington DC to Orlando, Florida. Season 1 develops around the first months after the moving, with Eddie as the focus of the story. He and his two younger brothers – Evan and Emery – struggle to fit in their new school, while his mother Jessica copes with a culture clash between her Taiwanese roots and a Florida

community with no Asian population, and his father Louis struggles to fulfil his American dream of owning a successful steakhouse restaurant. Season 2 storyline moves the spotlight to the entire Huang family, with more episodes revolving around Eddie's parents and introducing new characters, as they continue on settling in Orlando.

Figure 4.4. Still image of the five main characters in *Fresh off the Boat*



This TV program was found to be particularly appropriate for various reasons related to both format and content. First, the length of the episodes was adequate for a 1-hour lesson (the average running time of each episode was 21 minutes, including the opening credits). Secondly, it was a sitcom, a genre with which participants were familiar through watching similar TV programs – such as *Modern Family* or *The Big Bang Theory*, widely broadcasted in Spanish television. Third, it was serial in nature: although each episode developed a short story on its own, there was an overall storyline that linked the episodes to each other, which allowed participants to gather information about the characters as they continued on watching new episodes. Fourth, its content was appropriate for this particular age group, as it did not contain scenes with violence, sex, drinking, drugs or consumerism – without being too infantile. This was a key factor, since it had to be used

in a classroom setting. Fifth, it was engaging: the irony and humour of the series, as well as the 12-year-old rebellious yet innocent main character Eddie, were the perfect combination that made participants feel identified with him and get hooked into the story. Sixth, it was not strongly accented. Finally, at the time the intervention took place *Fresh off the Boat* had not been aired in Spain. This minimized the possibility that participants had watched any of the episodes before, which might have given them an advantage on the episodes' vocabulary and comprehension tests.

#### **4.3.2. Selection of the episodes**

The scripts of the first two seasons of the TV series were obtained from the website *Springfield! Springfield!* (URL: <https://www.springfieldspringfield.co.uk/>), a free-access database of TV and movie scripts available for research and educational purposes. Scripts were checked for spelling mistakes, and it was made sure that orthographic representations of proper nouns and marginal words were homogeneous, before proceeding to further analysis.

Each script was analysed using the online tool VP-Compleat from Lexical Tutor (Cobb, 2008; URL: <https://lxtutor.ca/>). This vocabulary profiler breaks down a text by word frequency in corpora, and provides information on how many words a text contains from each word frequency band. Since a main interest of the study was to assess vocabulary acquisition, it was decided that, ideally, target items had to be (1) potentially unknown – and therefore, not from the first frequency bands –, and (2) appear two or more times within each episode. The VP-Compleat analysis provided an overview of each episode's vocabulary, and allowed a first selection of potential target items based on frequency of occurrence and word family.

From the first and second season of *Fresh off the Boat*, 24 episodes were finally selected for the treatment – 11 episodes from season 1, and 13 episodes from season 2. Three non-consecutive episodes were skipped because either they did not contain enough potentially unknown lexical items to teach, or because schoolteachers considered they included an inappropriate topic or scene for 13-year-olds. Before excluding the episodes, it was checked that they did not include key information about the general storyline development, and therefore that its omission did not hinder comprehension of consecutive episodes. By the end of the intervention, participants had watched a total of 515 minutes of audio-visual input (8 hours and 35 minutes).

#### **4.3.3. Episodes' lexical profile**

It has been suggested that lexical coverage – the percentage of words known in a given input – might be one of the main factors, though not the only one, in determining comprehension (Webb, 2010a; Webb & Rodgers, 2009b). The level of difficulty of input has to be suited to learners' proficiency level (King, 2002; Rodgers & Webb, 2009b), since learning materials with an appropriate level ease students' possible frustration, and sufficient lexical coverage facilitates dealing with unknown vocabulary (Rodgers & Webb, 2010). Although an adequate lexical coverage does not ensure that discourse will be understood, it provides a valuable measure to estimate the vocabulary size needed to understand input, and therefore helps determine the series overall adequacy to a target population. The 24 episodes selected for the intervention were further analysed using the Range software (Nation & Heatley, 2002). As explained by Nation and Webb (2011):

“Research investigating the vocabulary size needed for comprehension has typically counted words using word families. The Range program provides the distribution of tokens, types and word families in each word lists. The percentage of tokens in each word list shows

the relative importance of knowing items from those lists. This percentage can be added up to determine the cumulative coverage. This reveals the vocabulary size necessary to reach the certain coverage points associated with comprehension and incidental learning". (Nation & Webb, 2011: 162-163).

The analysis of the lexical profile showed that, overall, the series reached 93.84% coverage at the 2,000 word-level and 95.70% coverage at the 3,000 word-level plus proper nouns and marginal words. Research on informal listening has suggested that a coverage of 90 – 95% might be enough (Noreillie, et al., 2018; Van Zeeland & Schmitt, 2013), so the series was considered adequate: participants in the present study had a mean vocabulary size of almost 2,000 words (which for this series represented a coverage of around 94%) and they had the additional support of the on-screen text, which ensured that input was challenging enough to promote learning but not overwhelming (Krashen, 2003). According to Rodgers and Webb's (2009b) corpus study, the mean lexical coverage of comedy is 93.99% at the 2,000 word level, and 95.64% at the 3,000 word level (Rodgers & Webb, 2009b: 419), which makes this series a typical example of the genre.

The lexical profile for the 24 episodes included in the intervention is shown in Table 4.1. The table displays the accumulative lexical coverage (in percentage) of the episodes at the 1,000, 2,000 and 3,000 word families. These estimates include proper nouns and marginal words. Further details on the number and percentage of tokens in each frequency band can be found in Appendix A.1. The addition of proper nouns and marginal words was based on findings from previous research in the area: Nation (2006) suggested that proper nouns have a minimal learning burden; and Webb and Rodgers (2009b) in their study of the lexical coverage of movies showed that if learners knew proper nouns and marginal words (e.g., ahh, ohh, huh) they could reach 95.76% coverage with the most frequent 3,000 word families. In the present study, proper nouns

make up 3.11% of the running words, adding more coverage than words from the 3,000 word level (1.62%) (see Appendix A.1). Considering that characters and locations reoccur throughout the episodes, it seems safe to assume learners were familiar with most of the proper nouns. By the end of the intervention, participants had been exposed to 69,350 tokens.

Table 4.1. Accumulated lexical coverage per episode

<b>Episode</b>	<b>1k %</b>	<b>2k %</b>	<b>3k %</b>
1	92.03	95.22	96.69
2	90.93	94.79	96.28
3	89.11	93.40	95.75
4	90.37	94.76	96.52
5	89.34	93.81	95.30
6	89.62	93.95	96.12
7	90.86	94.68	96.25
8	91.28	94.31	96.08
9	89.23	93.61	95.18
10	89.54	93.33	95.58
11	89.36	93.71	95.12
12	89.55	94.96	96.38
13	88.61	93.67	95.02
14	89.50	93.14	95.02
15	91.07	94.94	96.31
16	88.01	92.42	94.86*
17	91.52	94.99	96.39
18	87.74	91.52	94.60*
19	89.98	93.24	95.01
20	88.95	94.73	96.24
21	88.17	93.49	95.45
22	90.92	94.60	96.73
23	87.74	91.51	93.98*
24	88.84	93.29	95.95
<b>MEAN</b>	89.68	93.84	95.70

As can be observed, all episodes reach the 95% threshold at the 3,000 word level except for three (in sessions 16, 18 and 23, marked with a \*). The coverage provided by the first 3,000 words of the BNC/COCA word lists ranged from 93.98% to 96.73% between the episodes. Although the difference seems small (2.75%), research has shown that even a small increase in lexical coverage can already be beneficial for comprehension (Laufer & Ravenhorst-Kalovski, 2010), and since it could not be assumed that all episodes were equally difficult, the percentage of lexical coverage per episode was also included as an episode-related factor in Study 1, concerned with learners' comprehension of the episodes (see Chapter 5).

#### **4.3.4. Preparation of captions and subtitles**

Once it was established which episodes and lexical target items (see section 4.3.5 below) were to be included as part of the treatment, the next step was to edit and revise the two versions of the on-screen text support – captions and subtitles – to ensure that both language conditions were as similar as possible. English captions were available from the original DVD, but that was not the case of Spanish subtitles. The Spanish subtitles for the 24 episodes were downloaded from the website Open Subtitle, a free online search tool (URL: <http://opensubtitles.org>). To be able to edit them, the English captions were also downloaded from the same website and compared against the original captions in the DVD.

An editable version of both English captions and Spanish subtitles were created and edited using Aegisub, a free open-source subtitle editor (URL: <http://www.ageisub.org>). The resulting files were then revised so that each caption/subtitle complied with the basic guidelines of good subtitling practises. Each caption/subtitle appeared in the centre and bottom area of the screen, it had two lines maximum and it did not exceed the



recommended 37 characters per line for a TV program (Díaz-Cintas & Remael, 2014). Following the widely accepted 6-second rule, the longest possible caption/subtitle of two lines was shown on-screen no longer than 6 seconds (e.g., Bravo, 2005; Gielen & d'Ydewalle, 1989). Captions in English were verbatim, although elements of audio-description (e.g., [eerie music]) were removed to match the Spanish version. In some cases, because of the physical space available, the pace of soundtrack dialogue and the time limit, the text in Spanish had to be condensed (in no case such cases affected the TIs or their context). A common practise is, for example, to change the compound past tense in the dialogue to simple past tense in the on-screen text, such as changing 'I have done that' for 'I did that' (Koolstra, Peeters & Spinhof, 2002). The use of the Aeguisub editing tool also allowed maintaining the same text styling (typography, size and colour) across the two language conditions.

Each pair of lines of captions and subtitles was also compared to ensure that the number of encounters with the target items was identical in both the English and Spanish versions. When an English word had two translations in Spanish, one translation was consistently chosen over the other. For example, the English word *neighbourhood* appeared in the Spanish subtitles as both "barrio" and its synonym "vecindario"; in that case, it was decided to maintain the more common translation, as shown in Figure 4.5. The Spanish equivalent chosen per each target item was the same that would appear afterwards in the post-viewing task (see section 4.4.2.2).

Figure 4.5. Checking encounters with TIs across language conditions

Caption	Subtitle
It's a chance to forget that we're a lower-middle-class <b>neighbourhood</b>	Es una oportunidad para olvidar que somos un <b>barrio</b> de clase media baja
I just bought all the eggs from every store in the <b>neighborhood</b>	Acabo de comprar todos los huevos de todas las tiendas del <del>vecindario</del> <b>barrio</b>

### 4.3.5. Selection of target items

From each episode, five target items (TIs) were selected to investigate vocabulary acquisition through the TV series, making up a total of 120 TIs (5 TIs x 24 episodes). Three distractors (non-target items, NTIs) were also selected per each episode, and appeared in the post-viewing task<sup>10</sup> (see below). Prior knowledge of the TIs was assessed by means of three sets of pre- and post-tests (see section 4.3.2.2), administered at the beginning and the end of each term (eight episodes and 40 TIs per term). The TIs were only taught in the focused instruction groups (CF and SF), but all the four experimental groups were tested on the same 120 target items. TIs and NTIs were chosen according to frequency of occurrence within the episode (they had to appear a minimum of two times), and the low likelihood to be known by participants at this level of proficiency. A preliminary list of potential target items was elaborated and presented to the schoolteachers, who revised the selection and pointed out those TIs they considered likely to be known by their students, which were then replaced. The final

---

<sup>10</sup> Initially, post-viewing tasks did only include the five TIs from the episode. However, during the first viewing sessions – when participants started to get used to the sessions structure – it was observed that a couple of students in the focused condition were only trying to memorize the TIs. To prevent students from making a direct connection between the two tasks, NTIs were added in the post-viewing task from the 4<sup>th</sup> session onwards and, additionally, in the pre-task for T3.

selection of 120 TIs can be found in Table 4.2 below. Detailed information about the TIs and NTIs can be found in Appendix A.2 and A.3, respectively.

Table 4.2. Distribution of target items by word families in the BNC/COCA word list

	Target items			N (%)
	T1	T2	T3	
1,000	to fire, hell, to move	bill, to date, deal, to ground, to join, play, straight, to trade	to bet, enough, friendship, to share, tax, to trust, wall	18 (15.0%)
2,000	cheers, to handle, to hire, ride, to struggle, tough	to cash, to coach, cool, dump, fee, fridge, grade, neighbourhood, proud, ride, scary, to trip, van, weird	bakery, binder, to dare, to rush, threat	25 (20.8%)
3,000	curly, gross, nightmare, principal, shield	to bounce, bow, drill, guest	Christmas Eve, dull, envelope, to gamble, hen, trap, oven, raw, to tease, turkey	19 (15.8%)
4,000	Jewish, rib	costume, floor mat, jealous, rug, whale	review, wealth	9 (7.5%)
5,000	to ace, fake, franchise, to quit	cabbage	bunk, fireworks	7 (5.8%)
6,000	hedgehog	chess, neckless	haircut	4 (3.3%)
7,000	mall, napkin	kilt	oyster	4 (3.3%)
8,000	-	-	noodle	1 (0.8%)
9,000	billboard, buckle, to gloat, to loiter	-	orchids	5 (4.2%)
10,000	jukebox	teamwork	-	2 (1.7%)
11,000	-	prank	cranberry	2 (1.7%)
15,000	cushy	-	-	1 (0.8%)
16,000	janitor	piccolo	nerd	3 (2.5%)
18,000	crouton	-	-	1 (0.8%)
19,000	realtor	-	-	1 (0.8%)
Off-list	a/c, carpool, to figure out, to fit in, hairdryer, to hand out, knockoff, real estate	chickenpox, to hang out, stuffed animal	to cut off, flight attendant, fortune teller, to make sense, to mess up, to pick on, sleepover	18 (15.0%)

Target items belonged to different parts of speech, with the majority of them being nouns (59.2%), verbs (including phrasal verbs) (26.7%) and adjectives (11.7%). The remaining 2.5% were a small sample of adverbs, interjections and multi-word units (e.g., *Christmas Eve*). The 120 target items were from the first to the nineteenth frequency bands of the BNC/COCA word lists (Nation, 2012). Around 52% belonged to the first three most frequent word families (15% to the 1k, 20.8% to the 2k, and 15.8% to the 3k), 27% to the 4-10k, and 7% to the 11-19k, while the remaining 15% were off-list.

Frequency of occurrence of the TIs within each episode was between 2 and 14 times, with a mean frequency of 4.46 (SD = 2.5). Frequency within the target episode (i.e. the episode where the TI was taught to the focused groups) was used to select the TIs – as well as the NTIs. However, between the pre- and post-test learners could encounter the TIs in other episodes. To account for these other encounters, a more adjusted measure of frequency was used: frequency of occurrence within the term<sup>11</sup>. Frequency of occurrence within the term ranged between 2 and 20 occurrences, with a mean frequency of 5.55 (SD 3.5). Table 4.3 shows the minimum, maximum and mean number of encounters with the TIs per each term and the average across the year. Overall, 59% of the TIs were encountered between 2 and 5 times, 31% between 6 and 10 times, and 10% between 11 and 20 times.

---

<sup>11</sup> Some of the target items also appeared in terms in which they were not tested. For example, a target item taught during T3 might have already been encountered in episodes in T1 or T2. These prior encounters, however, were not taken into account. If learners had learnt the target item from prior terms, it would have been shown in the pre-test of the term. Detailed information on accumulated frequency of occurrence within each term can be found in Appendix A.2.

Table 4.3. Frequency of occurrence of TIs per episode and term

	n	Frequency within episode				Frequency within term			
		Min.	Max.	Mean	SD	Min.	Max.	Mean	SD
<b>T1</b>	40	2	11	4,02	2,465	2	20	5,45	3,843
<b>T2</b>	40	2	12	4,88	2,672	2	19	5,85	3,704
<b>T3</b>	40	2	14	4,48	2,342	2	15	5,35	3,043
<b>Total</b>	120	2	14	4,46	2,500	2	20	5,55	3,524

The software *Sketch Engine* (URL: <https://www.sketchengine.eu/>) was used to check the context of appearance of the TIs across the intervention. Sketch Engine is a widely used corpus management and query system that allows uploading your own sets of texts and analysis concordance and appearances in different parts of speech. That allowed to control for polysemy (e.g., *to move* as relocate) and take into account encounters with separable phrasal verbs (e.g., *cut off*, *cut it off*, *cut people off*). Figure 4.6 shows an example of the output provided by the Sketch Engine tool.

Figure 4.6. Sketch Engine output sample

S02E14.docx	Emery, who I'm sure got preside...	Flight attendant	? I like travel, snacks, and being helpful. It
S02E14.docx	, and being helpful. It checks all the boxes. Oh,	flight attendants	don't even get a seat on the plane. They walk back
S02E14.docx	stuff, mom. Oh, that's okay. That's for being a	flight attendant	, not for tennis. Don't be silly. I'm just
S02E14.docx	tennis. We could do father-son tournaments.	Flight attendant	, my ass. He'll get sports scholarships. I'll be
S02E14.docx	your father so you can achieve excellence. A	flight attendant	wouldn't have the eggs to do that. Thanks, mom.
S02E14.docx	I want to do, like flying around the world as a	flight attendant	. Well, I guess we need to embrace the kid we have.
S02E14.docx	. Absolutely. Space camp? Astronauts are like	flight attendants	, but for science. Travel, snacks, helping
S02E14.docx	? Embracing the kid we have. If he's gonna be a	flight attendant	, he's gonna be the best kind. I'm gonna float

## 4.4. Testing instruments

### 4.4.1. Language tests

#### 4.4.1.1. *Assessing general proficiency: Oxford Placement Test*

The Oxford Placement Test (OPT) was used as a general measure of proficiency because (1) it includes a listening section, which was deemed appropriate considering that listening skills are especially relevant in this learning environment; (2) it takes 60 minutes to complete, which makes it suitable for a 1h-class; (3) it has multiple-choice format, so students do not have to write; (4) it is easy to score; and (5) it is a validated test suited from levels A1 to C2+, which would allow comparing the proficiency level of participants in the present study with participants in other studies and different populations.

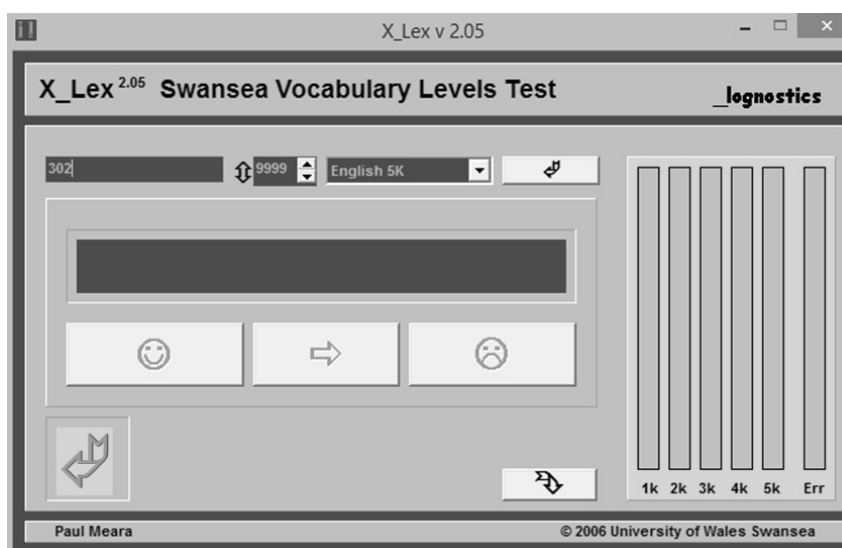
The OPT is an easy-to-administer general proficiency test calibrated against the levels of the Common European Framework of Reference for Languages. The test is divided into two main sections – a Listening Test and a Grammar Test – consisting of a 100 items each (200 items in total). Part 1 is the Listening Test, which takes 10 minutes, test-takers hear each sentence (= item) once and they have to choose between two possible answers, which are both possible and grammatically correct; thus, the correct answer can only be selected by understanding *aurally* what is being said (it cannot be deduced from the context). Part 2, the Grammar Test, takes 50 minutes to complete and it is a written multiple-choice test. Three options are given to complete each sentence. The items are contextualized, so grammar is tested in relation to meaning and it involves reading comprehension as well as language knowledge.

#### 4.4.1.2. Assessing vocabulary size: X-Lex test

The X\_Lex test (Meara & Milton, 2003) was used to assess learners' approximate vocabulary size. This test was chosen because it is simple in format and it can be administered in a short time – it takes approximately 10 minutes to complete – while allowing testing for a large number of items (Milton, 2010).

The X\_Lex test (which has currently been replaced by its updated version, the Yes/No Levels Test) is a computerized test in check list format that measures learners' L2 receptive vocabulary knowledge of the most frequent 5,000 words in the language, and estimates overall knowledge of this vocabulary (Milton, 2010). The test randomly presents 120 items selected from the first five frequency bands, plus a number of non-words designed to look and sound like words in the target language – in this case English. In the test, words are presented one at a time on the computer screen and participants have to indicate whether they know the word or not by clicking on the appropriate button (a smiley face ☺ for “yes” and a sad face ☹ for “no”). In case test-takers do not recognize the word or believe it does not exist, the test instructs them to select the option “no”. Figure 4.7 below shows how the test looks like. The amount of “yes” responses to real words gives a preliminary estimate of the words known, while the number of “yes” responses to non-words allows adjusting for guessing and overestimation of knowledge (Milton, 2010). The test provides an overall final score of words known up to the 5,000-word level.

Figure 4.7. X\_Lex vocabulary size test



Because of its brevity, it might be argued that the X\_Lex test is less reliable than other more detailed tests (Milton, 2010), such as Nation's (2001) Vocabulary Levels Test (VLT). However, the VLT format – which includes a multi-item matching task per each word – is more complex, and relies not only on the test-taker's knowledge of the target words but also on the knowledge of the words in the definitions. Considering the age and expected proficiency of participants in the present study, it was considered that the X\_Lex test was a better alternative. The X\_Lex test has also been shown to correlate strongly with proficiency (Meara & Jones, 1988; Miralpeix, 2009). In a validation study, Milton and Meara (2003) tested students at every level of the CEFR and estimated their vocabulary size using the X\_Lex test. Results showed a vocabulary size of 1,500 to 2,500 words in the X\_Lex corresponded to the A2 level.



#### *4.4.1.3. Vocabulary pre-test and post-tests*

To assess participants' knowledge of the target vocabulary items before and after the treatment, a pre- and post-test design was used. As stated before, because of school calendar constraints, the intervention itself had to be divided into 3 terms of 8 viewing sessions each, and therefore, participants completed 3 sets of pre-/post-tests, and in each one they were tested on the 40 TIs corresponding to that term. Pre-tests were administered 1 or 2 weeks prior to the first viewing session to reduce pre-test effects. The decision of having 3 sets of tests was also made to avoid decay due to having the post-test too far from the first sessions. Tests were paper-and-pencil based.

The vocabulary tests consisted of two parts: (1) a form transcription test, and (2) a meaning recall test. Participants first listened to each TI twice – with a 5 seconds pause between repetitions – and had to write down the English word (word-form transcription). Then, they had to translate the item into Catalan or Spanish, or provide a short definition (word-meaning recall). There was a 10-second break between each different item to allow participants to write down translations, and 3-5 extra minutes were given at the end of the test so participants could fill in any missing answers. The first part provided the measure for gains in form, while the second one measured gains in meaning. This type of test was chosen because, in order to assess the benefits of captions and subtitles for vocabulary learning, tests had to be congruent with the input-modality (Mohd Jelani & Boers, 2018) – written L2 word prompts in the test could have given an advantage to the captions groups.

The pre- and post-test were always administered by the researcher. Participants were explained how they had to complete the test and were told to remain silent throughout the task. At the beginning of the test, there were five trial items, so participants could get familiar with the format, assess the volume of the audio and ask any questions they had

before starting. Target items were recorded by two American native speakers, one male and one female. They were asked to pronounce the words clearly but at a normal rate, and to try to maintain a neutral American accent (which was the accent of the TV program chosen for the intervention). The order of appearance of the target words was randomized using an online randomizer tool (<http://www.random.org>), and the final order was revised so that similar sounding words (i.e. “rug” and “raw”) did not appear consecutively.

The pre and post-tests were identical, and the format was maintained across the three terms as well (only changing the target items included). Figure 4.8 provides an example of the format of the pre- and post-tests. A sample of the vocabulary tests can be found in Appendix A.4.

Figure 4.8. Vocabulary test format



One (1-second pause) target item (5-second pause) target item (10-second pause)  
 Two (1-second pause) target item (5-second pause) target item (10-second pause)

	ENGLISH	CATALÀ / CASTELLANO
1		
2		
3		
4		

#### 4.4.1.4. Delayed vocabulary post-tests

Eight months after the intervention finished, a delayed post-test was administered to assess the retention rates of the vocabulary learnt during the last term of the intervention. This post-test was identical in format and content to the post-test in T3. Therefore, it

assessed the retention of the 40 TIs appearing in the last eight episodes of the intervention. (i.e. the 40 TIs tested in T3).

A second delayed post-test was also administered to assess partial knowledge of the TIs. A first analysis of the vocabulary gains data from T1 and T2 had indicated that the test format might have been too demanding. The relative low meaning recall scores attained across experimental groups (see Study 2, Chapter 6) suggested that the test might have fallen short in detecting participants' actual knowledge of the TIs. It was possible that learners did not recognize a word aurally, but that they could produce the meaning if they saw it. In view of this, a second meaning recall test was designed including the written forms of the TIs. In this case, participants only had to write down the translation or short definition, as shown in Figure 4.9 below. Besides the fact that it provided the written form of the TIs, the layout of the test was identical to the other vocabulary tests.

*Figure 4.9. Vocabulary delayed post-test (written format)*

	<b>ENGLISH</b>	<b>CATALÀ / CASTELLANO</b>
<b>1</b>	fake	
<b>2</b>	jukebox	
<b>3</b>	napkin	
<b>4</b>	chickenpox	

## 4.4.2. Intervention tests

### 4.4.2.1. Vocabulary pre-viewing tasks

As a means to direct learners' attention to the TIs and provide a first contact with the form and meaning of the words, focused groups (CF and SF) started each session with a pre-viewing activity. In this task, participants were taught the five TIs – plus three NTIs – appearing in the episode. These vocabulary tasks included word searches, crosswords, fill-in-the-gaps, matching words and definitions and matching words and images (see below)<sup>12</sup>. A sample of each type of pre-viewing activity can be found in Appendix A.5. Participants were given 5 minutes to complete the task individually, which was later corrected orally by the teacher and collected before viewing the episode. Each of the five types of activity used to draw attention to the target vocabulary in the pre-task was roughly used the same number of times. Since this was a longitudinal study, it was decided to use a variety of teaching activities to avoid excessive repetition and potential boredom. This also made the activities more similar to what they do normally in the regular classroom. To avoid favouring one language condition over another, the definitions or context accompanying these activities were provided alternately in English and Spanish. Definitions in English were adapted to the proficiency level of the sample, and were revised by an English native speaker. Table 4.4 shows the distribution of types of task and language use across the 24 episodes.

---

<sup>12</sup> Samples of the testing instruments (including different models of pre-task activities) were run with two learners of the same age and similar proficiency to spot potential problems, especially regarding the pre- and post-tasks, as well as the comprehension tasks.

Table 4.4. Distribution of pre-viewing tasks by type and support language

	Number of pre-tasks in which this type of task was used	ENGLISH definition or context available	SPANISH translation, definition or context available
Matching word-image	6	6	0
Word search	5	1	4
Fill-in-blanks	6	1	5
Crossword	7	4	3
<b>TOTAL</b>	<b>24</b>	<b>12</b>	<b>12</b>


In the matching word-image task, learners were presented with sentences in English where the TIs were used in context, and they had to match each sentence with the corresponding image, as shown in Figure 4.10.

Figure 4.10. Matching task sample


**OVEN**  
*I just put the pizza in the **oven**; it will be ready to eat in 15 minutes!*

**THANKSGIVING**  
*For Thanksgiving, families have lunch together and eat delicious food.*

1



2



In the word search task, learners were presented with the definitions of the TIs – in English or Spanish and had to find the corresponding word in a word search. The first letter of the word was provided next to the definition, and it was also circled in the word search (an extra letter was also circled as a distractor). Figure 4.11 provides an example of such type of task.

Figure 4.11. Word-search task sample

<p>1. An agreement between two or more people. (4 letters) <b>D</b> _____</p> <p>2. A vehicle that is larger than a car and that is used for transporting people or things. (3 letters) <b>V</b> _____</p>	<pre> K (V) A N A E V W X E I M Q T L Q U F A M A G F T H J E V Q B N K (D) S O L </pre>
--	--

In the fill-in-the-blanks task, learners were presented with sentences with blank spaces, and they had to complete them using the words given. Translations or short definitions of the TIs were also provided in English or Spanish, as shown in Figure 4.12.

Figure 4.12. Fill-in-the-blanks task sample

<p>Please, could you _____ the worksheets to your classmates?</p> <p>Sometimes I have _____ when I'm sleeping.</p> <p>My favorite food is _____ with roasted vegetables.</p> <p><b>To hand out:</b> repartir</p> <p><b>Ribs:</b> costillas</p> <p><b>Nightmares:</b> pesadillas</p>
---

Finally, in the crossword task learners were presented with definitions of the TIs in English or Spanish. They had to match them with the TIs provided, and then fill in the crossword, as shown below in Figure 4.13.

Figure 4.13. Crossword task sample

1. Reírse o meterse con alguien con intención de hacer daño.

3. Persona cuyo trabajo es ayudar a los pasajeros que están viajando en un avión

FLIGHT ATTENDANT

PICK ON

#### 4.4.2.2. Vocabulary post-viewing tasks

After viewing the episode, all groups were given an immediate post-viewing vocabulary task, intended to ensure that learners were paying attention to the viewing of the episode. The tasks included the five TIs taught in the two focused groups plus three distractors, different from those appearing in the pre-task. Participants heard the eight items twice, they had to write them down and select the correct translation amongst the options provided (see Figure 4.14). This task was not corrected, but it was collected before the following activity (a comprehension task) was given. As in the case of the pre- and post-tests, target words were recorded by the same two American native speakers. A sample of the post-viewing task can be found in Appendix A.6.

Figure 4.14. Post-viewing task format

<b>1)</b> _____	
a) esponjoso	d) exigente
b) acomodado	e) hinchado
c) pobre	f) No lo sé
<b>2)</b> _____	
a) arrebatarse	d) lamentarse
b) flotar	e) deleitarse
c) brillar	f) No lo sé
<b>3)</b> _____	
a) orgulloso	d) alto
b) avergonzado	e) nublado
c) entusiasta	f) No lo sé

The post-viewing activity was a meaning recognition task. The format was a multiple-choice with six options, including the correct translation, four distractors and an “I don’t know” option. Besides the correct answer, the four distractors were always selected using the same criteria: one distractor was a word that sounded similar to the TI, one had a similar meaning, one was a hapax (a word appearing in the episode but with no similitude in terms of form nor meaning), and the fourth one was either a phonetic or a semantic distractor again. Figure 4.15 shows how the multiple-choice options were created (left) and an example (right).

Figure 4.15. Criteria for elaborating the distractors in the post-viewing task

<b>Target item</b>	<b>Fridge</b>
a) correct translation	a) nevera ( <i>fridge</i> )
b) distractor (similar phonetically)	b) puente ( <i>bridge</i> )
c) distractor (similar semantically)	c) congelador ( <i>freezer</i> )
d) distractor (b or c)	d) pescado ( <i>fish</i> )
e) distractor (hapax)	e) etiqueta ( <i>label</i> )
f) <i>I don't know</i> option	f) No lo sé ( <i>I don't know</i> )



#### 4.4.2.3. Content comprehension tests

Comprehension was assessed by means of a post-viewing task administered at the end of each viewing session, and participants could not pre-view the test. Each test consisted of 10 items, including five multiple-choice items (MC) and five true/false items (TF) (see Figure 4.16, and see Appendix A.7 for a sample of the comprehension test). Using a variety of question types provides a more balanced assessment (Buck, 2001) and participants were already familiar with these two item formats. Additionally, both provide a quick and reliable method for testing understanding of the content. MC items had three options (one correct and two distractors). All items were designed in a way that the information given by the image alone was not sufficient to answer the question, and the two distractors in the MC items did not provide clues to respond to other questions.

Figure 4.16. Examples of a MC (top) and TF (bottom) format

Why did the Huang family not go to Washington D.C.?

- a) Because the flight was cancelled.
- b) Because they got the date wrong.
- c) Because they slept in and missed their flight.

**V / F** Eddie is sad because he cannot visit his family in Washington for New Year.

Comprehension items also included two types of questions: textually explicit items (TE) (when the information is explicitly stated in the text, and it could be underlined in the script), and inferential items (IN) (when the information is found by combining or deducing from different pieces of information, integrating them to understand the central gist or idea). The operationalization of item type was based on an adaptation from Davey

and McBride (1986), Alptekin and Erçetin (2010), and Rodgers (2013). Figure 4.17 provides an example of the two item types, in MC and TF format, including the script fragment where the answer can be found.

Figure 4.17. Examples of TE and IN formats

**Textually explicit items (TE), in MC and TF format**

With what does Eddie trip over when he breaks his arm?

- a) With a mechanical bull.
- b) With a white rug.
- c) **With a cord.**

SCRIPT: “Okay! Hey, boys, we can’t let mom know **Eddie tripped over the cord of the mechanical bull and broke his arm.** (...) Now tell these nice people the truth about how you broke your arm. Eddie. You were at Cattleman’s **and you tripped... On the cord to the new mechanical bull.**”

V / F Neither Jessica nor Louis knew that to be a relator you needed a licence.

SCRIPT: “[Jessica] apparently, you need a license to sell houses! Did you know this?! [Louis] **I did know this. I told you this, and you ignored me.**”

**Inferential items (TE), in MC and TF format**

Why do Jessica and Louis ask Nicole to babysit their children?

- a) Because Eddie asked for it to be able to talk to her.
- b) Because their grandma is too old to take care of the children.
- c) **Because Nicole is the cheapest option.**

SCRIPT: “[Eddie] Babysitting? Why can’t grandma watch us? [Jessica] Because last time, she charged me \$100 an hour. All right, we’ve got to get going. Nicole, some ground rules... no scary movies.”

V / F Evan has a sever gambling problem

SCRIPT: “[Eddie:] I can’t believe you lied to me. I thought you were getting picked on. [Evan:] I’m addicted, okay? I’m addicted to playing Pogs. I don’t even get what you see in that noise. I love the thrill of taking risks, laying it all on the line.”

A second researcher checked the classification of the items into TE or IN type (a detailed guideline on what was considered one type or another was provided, along with the scripts of the episodes). Disagreements were discussed, and appropriate changes were made when an item was found to be problematic. Table 4.5 shows the final distribution of the total 240 comprehension items (10 items x 24 tests).

*Table 4.5. Comprehension items by format and type of question*

	<b>Textually explicit</b>	<b>Inferential</b>	<b>Total</b>
<b>Multiple-choice</b>	59	61	120
<b>True/False</b>	60	60	120
<b>Total</b>	119	121	240

Tests were administered in Spanish as the main purpose was to assess learners' content comprehension, and the use of the L1 ensured avoiding errors attributable to poor comprehension of the questions (Vandergrift, 2007). Because of this, textually explicit questions were formulated using paraphrases and synonyms rather than literal excerpts from the audio, since this might have prompted learners in the subtitles group just to choose the options containing the vocabulary they were seeing in the subtitles (Taylor, 2005). This was done to avoid lexical overlap, which tends to be a predictor of easy items, since test-takers tend to select options that contain vocabulary that they recognize from the input (Buck, 2001). For this reason, the term "textually explicit items" is preferred over "literal items".

### 4.4.3. Questionnaires

Four questionnaires were administered at three different points of the intervention (i.e. pre-intervention, post-intervention and delayed post-intervention) to collect data on students' preferences, attitudes and perceptions – and assess whether they had changed over time. The first two questionnaires were given prior to the intervention (in September 2016), together with the initial proficiency tests. The third questionnaire was given just after finishing the intervention (in June 2017), and the fourth one was administered eight months after the end of the intervention (in February 2018). Questionnaires were pen-and-paper based and completed in class under the supervision of the researcher and the teachers. All of them were provided in Catalan and Spanish. A sample of each of the four questionnaires can be found in Appendix A.8, A.9, A.10 and A.11<sup>13</sup>. Data from the questionnaires is discussed in detail in Study 4 (Chapter 8), and it will also be included in Study 1 (Chapter 5).

#### 4.4.3.1. Pre-intervention questionnaires: participants' background

The first background questionnaire (BQ1) focused on participants' prior experience with viewing audio-visual input in English. It included a section on on-screen text preferences, frequency of exposure, and perception of usefulness of the on-screen text support.

---

<sup>13</sup> Questionnaires included a number of questions that will not be addressed in the present study (they appear in the appendices). For clarity purposes, only items analysed will be presented in this chapter.

Preference for or against on-screen text was assessed through responses to 10 statements – five in favour of on-screen text and five against it (BQ1#1)<sup>14</sup>. For each of the 10 statements, respondents had to rate their identification with them in a 5-point scale from 0 (“never”) to 4 (“always”). Verbal labels were included through the five categories, as see in Figure 4.18 below. Preference for or against on-screen text was determined by adding up the responses in each of the two sets of five statements, and the one with higher value taken to represent the students’ preference (e.g., if a student had a total of 15 points in the set in favour, and 4 in the set against, he or she was considered to have a preference *for* on-screen text). Respondents with the same amount of points (or a difference of less than 2 points between the two sets) were considered as having no preference for or against on-screen text. Note that this measure did not distinguish between L1 and L2 text, but between having text support or not.

Figure 4.18. Item #1 in BQ1

When I watch series or movies in English with subtitles...	Never	Rarely	Sometimes	Often	Always
I prefer to have subtitles.	0	1	2	3	4
Subtitles help me understand the dialogue.	0	1	2	3	4
I understand the plot better if I read the subtitles.	0	1	2	3	4
Subtitles help me recognize the words I hear.	0	1	2	3	4
I read the subtitles before listening to the voices.	0	1	2	3	4
I prefer to watch the series or movie without subtitles.	0	1	2	3	4
I concentrate more on the images than on the subtitles.	0	1	2	3	4
Subtitles distract me.	0	1	2	3	4
I concentrate more on the dialogue that I hear than on the subtitles..	0	1	2	3	4
I get tired reading the subtitles.	0	1	2	3	4

<sup>14</sup> This type of measure is similar to the one used to determine learners’ learning styles in Cohen, Oxford and Chi’s (2001) self-reported survey.

Frequency of viewing audio-visual input in English – with subtitles in Catalan or Spanish (L1), captions in English (L2) or without on-screen text – was assessed through three multiple-choice questions, one per each choice of on-screen text (BQ1#2). Students were asked with which frequency they normally watched TV series or films (including online videos, such as tutorials or YouTube videos). Six options were provided, from *never* to *more than 6 hours per week* (see Figure 4.19).

Figure 4.19. Item #2 in BQ1

With what frequency do you watch movies or TV series in English with...

... subtitles in Catalan or Spanish:

never    1-2h/year    1-2h/month    1-2h/week    2-6h/week    +6h/week

... captions in English:

never    1-2h/year    1-2h/month    1-2h/week    2-6h/week    +6h/week

... without subtitles:

The last item in the pre-intervention questionnaire focused on the perceived usefulness of having on-screen text for language learning purposes (BQ1#3). This was assessed using a 5-point Likert scale, from “not at all” to “a lot” (the scale only included verbal labels). As a follow-up, students were also asked to motivate their answer through an open-ended question, as shown below (Figure 4.20).

Figure 4.20. Item #3 in BQ1

Do you think subtitles help you learn English?

not at all    not much    a bit    quite a lot    a lot

Why? .....

The second background questionnaire (BQ2) assessed participants' exposure to and use of English outside the classroom environment. It was designed by the GRAL research group at the University of Barcelona, and it has been used in multiple studies from members of the research group (e.g., Muñoz, *forthcoming*; Muñoz, Cadierno & Casas, 2018). The questionnaire is divided into four main parts: (I) use of English outside the classroom (watching TV or movies, playing games, listening to music, reading, talking face-to-face, using the internet), (II) stays abroad, (III) language summer camps and (IV) extracurricular language classes. Data collected through this questionnaire was used to measure participants in the current sample against a much larger sample of respondents in the same context in terms of frequency of exposure to English outside the formal context (see Chapter 8).

#### *4.4.3.2. End-of-intervention questionnaire: assessing the intervention*

The third questionnaire (BQ3), administered at the end of the school year, consisted of two parts. The first part contained several items identical to those in BQ1, as the aim was to observe whether there had been changes in viewing habits or perceptions about captions and subtitles after the intervention. Questions regarding preference for or against on-screen text (BQ3#1), frequency of viewing OV input in English (BQ3#2) and perceived usefulness of having on-screen text for language learning purposes (BQ3#3) were identical to the corresponding items in BQ1 (see above). Additionally, students were asked in what subtitle language (L1 or L2) they normally preferred to watch input in English (BQ3#4). To delve into the reasons behind students' preferences, a list of options was provided for each of the three on-screen text choices – subtitles, captions or no-text. Respondents were instructed to tick as many options as they needed for each

language choice (or choose none if they did not watch OV input) (BQ3#5), as shown in Figure 4.21.

Figure 4.21. Items #4 and #5 in BQ3

**When you watch movies or TV programmes in English, you generally prefer...**

subtitles in Catalan or Spanish     subtitles in English     no subtitles     I never watch original versions in English

**For which reasons do you watch series or movies in English with or without subtitles?**  
*Select all the options you want in EACH column, and add extra ones if needed:*

Subtitles in Catalan/Spanish	Subtitles in English	Without subtitles
<input type="checkbox"/> I like to hear the original version	<input type="checkbox"/> I like to hear the original version	<input type="checkbox"/> I like to hear the original version
<input type="checkbox"/> There's no dubbed version	<input type="checkbox"/> There's no dubbed version	<input type="checkbox"/> There's no dubbed version
<input type="checkbox"/> There aren't English subtitles	<input type="checkbox"/> There aren't Catalan/Spanish subtitles	<input type="checkbox"/> There aren't subtitles available
<input type="checkbox"/> I don't like English subtitles	<input type="checkbox"/> I don't like Catalan/Spanish ones	<input type="checkbox"/> I don't like subtitles
<input type="checkbox"/> I need subtitles to understand	<input type="checkbox"/> I need subtitles to understand	<input type="checkbox"/> I don't need them to understand
<input type="checkbox"/> Subtitles help me understand better (but I don't need them)	<input type="checkbox"/> Subtitles help me understand better (but I don't need them)	<input type="checkbox"/> Subtitles distract me
<input type="checkbox"/> I watch series or movies with someone who wants/needs them	<input type="checkbox"/> I watch series or movies with someone who wants/needs them	<input type="checkbox"/> I watch series or movies with someone who doesn't want/need them
Other reasons:.....	Other reasons: .....	Other reasons: .....
.....	.....	.....
.....	.....	.....
.....	.....	.....

The second part of the questionnaire was designed to capture the students' overall feeling of learning from the intervention. The first two items asked learners how attentive they had been (BQ3#6), and whether they had overall enjoyed the episodes or had not (BQ3#7). They had to rate both aspects in a 5-point scale from "not at all" to "a lot", as shown in Figure 4.22.



Figure 4.22. Items #6 and #7 in BQ3

*In general, do you consider that you have been attentive during the viewing of Fresh off the Boat?*

not at all     not much     a bit     quite a lot     a lot

*In general, did you like the episodes?*

not at all     not much     a bit     quite a lot     a lot

The next two questions focused on students' feeling of learning from the intervention. Students had to select – from a list with several options – all of those they agreed with (i.e. check-all format, see Figure 4.23 below). The first question focused on aspects related directly to language learning (BQ3#8), while the second one included more general aspects related to learning (BQ3#9). The options provided were based on the most common comments elicited in the open-ended question about usefulness in BQ1 (BQ1#3), as well as during the individual interviews carried out at the end of T2.

Figure 4.23. Items #8 and #9 in BQ3

*In what aspects has the series helped you? (you can select more than one option)*

- Matching word & pronunciation
- Matching word & meaning
- Vocabulary
- Pronunciation
- Spelling
- Retention of new information
- General comprehension
- Listening in general

*From the following list, select ALL the options with which you agree*

- Now I understand the series better (the dialogues, the actors) than at the beginning.
- As a listening activity, this is more natural than other types of listening activities.
- I've learnt about cultural aspects (about the Chinese culture).
- This activity has motivated me.
- I felt relaxed during the sessions.
- I was paying less attention than in other activities because I was more relaxed.
- I have the feeling that I have learnt with these sessions.
- I would like to continue with a similar activity next year.

In the last question (BQ3#10), learners were asked to give an overall assessment to the one-year-long activity, in a 7-point scale from 0 to 6 – where 0 was the lowest mark and 6 the highest – as shown in Figure 4.24. They were given some extra space so they could also express their opinions or complain about the activity.

Figure 4.24. Item #10 in BQ3

*Overall rating of the activity:*

  0  1  2  3  4  5  6 

#### 4.4.3.3. Delayed questionnaire: long-term effects

A short questionnaire was administered together with the delayed post-test that participants completed 8 months after the intervention. It focused on three main aspects. The first question (BQ4#1) – which was identical to BQ1#2 – inquired about the current frequency of viewing OV – with captions, subtitles and without text. A second question asked participants whether they were watching more or less OV input than the previous year (BQ4#2). This was done with a yes/no format question, and respondents were asked to motivate their answers. A follow-up question asked to specify whether now they preferred watching input in English with captions, subtitles, or without on-screen text (BQ4#3), as shown below (Figure 4.25).

Figure 4.25. Items #2 and #3 in BQ4

**Do you watch more movies, TV series or videos in English this year compared to last year?**

yes  no    Why? .....

**If your answer is yes, you prefer...**  captions    subtitles    no-text    I don't watch anything

#### 4.4.4. Interviews

In order to gather more information on students' insights about the intervention, one-to-one interviews with a sub-sample of 17 students were also arranged by the end of the second term (around March 2017). Based on their OPT score, four participants were initially selected from each experimental condition: two high-proficient students and two low-proficient students (within the proficiency range of the sample). A fifth student was interviewed because one of the participants' interview lasted only 4 minutes – which was

deemed insufficient. While participants were pre-selected by the researcher, they participated voluntarily. Interviews were carried out individually, at the school, and lasted around 10 - 15 minutes. They were conducted in Catalan or Spanish – students were asked to choose the language they felt more comfortable with – and were audio-recorded with their permission. Interviews were semi-structured and addressed three main areas: (1) insights about the intervention, (2) learning English in general, and (3) watching audio-visual materials in English outside formal learning environments.

A second round of interviews was also carried out during the delayed-post-testing phase – eight months after the end of the intervention – with the same 17 participants, who agreed once more to be interviewed by the researcher. Again, they were carried out individually, in Catalan or Spanish, lasted around 5 - 10 minutes, and were audio-recorded. The focus of attention was on changes in attitudes and viewing habits since the intervention.

## **4.5. Procedure**

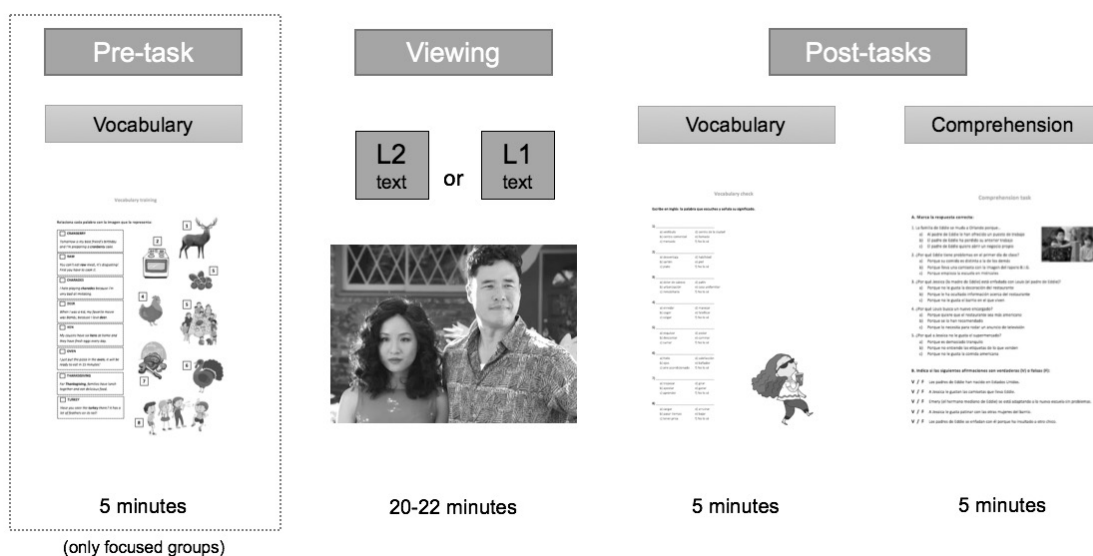
### **4.5.1. Viewing sessions**

All 24 viewing sessions followed the same procedure, with slight differences across the four classes depending on the experimental condition they had been assigned to. The two groups with specific focus on vocabulary (CF and SF) started each session with a pre-viewing task aimed at teaching the five TIs appearing in that episode. Pre-viewing activities included matching exercises, word searches, fill-in-the-blanks tasks and crosswords (see section 4.4.2.1). Students had around 5 minutes to complete the task individually, which was then corrected orally, and collected by the teacher before the screening of the episodes. Participants in the two non-focused instruction groups (CNF

and SNF) did not perform this pre-viewing task, and their sessions started directly with the viewing of the episode. Then, all four groups watched the corresponding 20-minute episode with either captions (CF and CNF) or subtitles (SF and SNF). The episode was displayed in a standard 4:3 projector screen at the front of the class.

After the viewing, all four groups (CF, SF, CNF and SNF) completed the same two post-viewing tasks, namely the vocabulary task and the content comprehension test, which were given to encourage learners to pay attention to both vocabulary and content (see section 4.4.2.2 and 4.4.2.3). After the episode, the teacher handed out first the vocabulary post-task – which included the five TIs plus another three distractors –, and students had 5 minutes to complete it before it was collected and the comprehension test was given. Students then had 5 minutes to complete this last task. Teachers reported that the time allowed for the completion of all tasks during the session was adequate and that students did not rush to finish the activities. The two post-task were not corrected by the teachers, and students did not get feedback on them. Figure 4.6 provides an overview of a viewing session.

Table 4.6. Structure of a viewing session



#### **4.5.2. Teachers' training**

Before the summer break prior to the intervention (July 2016), a meeting was scheduled with the head of studies, the head of the English department and the English teachers' team. During this meeting, a general overview of the project and a preliminary schedule was presented. Considering that it was a yearlong intervention and a high degree of commitment from the teachers was going to be required, it was deemed appropriate by both the researcher and the head of studies to take into consideration teachers' motivation. Since this meeting occurred before the teachers were assigned to their next-course classes, only teachers who were interested in taking part in the project were assigned to the targeted Grade 8 groups. A second meeting – only with the Grade 8 English teachers – was held at the beginning of September, prior to the start of the academic year. A calendar for the forthcoming year was established taking into account school activities and bank holidays, as well as other schedule concerns (such as exams periods). Teachers also received a more detailed explanation regarding the intervention, and were informed about the activities and structure of the viewing sessions.

Because the intervention was carried out within the regular EFL classes, the viewing sessions were therefore generally conducted by the schoolteachers themselves. Four teachers took part in the project, and were trained and supervised by the researcher before and during the intervention. A detailed rubric of the sessions' structure was provided to ensure that all sessions across the four experimental groups were run, as much as possible, in the same manner. The first two viewing sessions were conducted by the researcher in all four groups – with the teacher as an observer – to demonstrate how the session had to be run. The third viewing session was conducted by the teachers and overseen by the researcher, who made the appropriate corrections if deemed necessary. From the fourth viewing session onwards, teachers run the sessions by themselves. Exceptionally, the researcher carried out a session if a teacher was absent

(to avoid mismatches in the schedule amongst the four intervention groups). The fact that the viewing sessions always followed the same structure helped teachers but also students to quickly familiarize with the procedure.

With the exception of the comprehension test and the vocabulary tasks, all other tests –proficiency tests, vocabulary tests and background questionnaires – were always administered by the researcher. Interviews were also carried out by the researcher.

#### **4.5.3. Schedule of the intervention**

The pedagogical intervention took place during a complete academic year, starting in mid-September 2016 and finishing in June 2017, and it was integrated as part of their regular EFL lessons. As mentioned above, in Grade 8, students have four 1-hour long EFL lessons per week, and one of those lessons was dedicated to the intervention<sup>15</sup>. Therefore, viewing sessions were generally held once a week, but the schedule had to be adapted the school calendar. Table 4.8 below provides a detailed overview of the intervention's final calendar from September 2016 (Session 1) to June 2017 (Session 32).

At the beginning of the intervention, students completed the two parts of a general language proficiency test (Oxford Placement Test), the X\_Lex vocabulary test and the first two background questionnaires. The 24 viewing sessions were divided into 3 terms in order to accommodate them in the school calendar. At the beginning of each term, students were tested on their prior knowledge of the 40 target items corresponding to

---

<sup>15</sup> In the session chosen, the class was split into two groups, that is, only half of the students were in the classroom – while the other half was sitting in another subject class (e.g. half had English, half had social sciences). That meant that, with a reduced number of students in class, viewing conditions were better.

the eight episodes they would watch during the term, and they repeated the same test at the end of the term (e.g., pre-test T1 was the same as post-test T1). After the intervention, students completed the third questionnaire, and took part in an eye-tracking session<sup>16</sup>. The gap between sessions 12 and 13 was due to a 3-day local festivity; the one between sessions 13 and 14 corresponds to the Christmas break; the one between sessions 22 and 23 was due to a school trip (each class went on the trip on different days); and the one between sessions 25 and 26 corresponds to the Easter break.

In summary, the complete intervention extended over 9 months and included 33 sessions. The delayed post-test was administered 8 months after the end of the intervention. Interviews with a subgroup of students were conducted during the first two weeks of April, and follow-up interviews were carried out on the same week the delayed-post-test was administered, on February of the following year.

Table 4.7. Schedule of the intervention

SESSION	WEEK	CONTENT	Video running time
1	Week 1	Pre-test T1 Questionnaire BQ1 OPT Listening test Questionnaire BQ2	
2	Week 1	OPT Grammar test Vocabulary size test	
3	Week 2	Viewing session 1: Season 1, Episode 1 - <i>Pilot</i>	22
4	Week 3	Viewing session 2: Season 1, Episode 2 - <i>Home Sweet Home-School</i>	21
5	Week 4	Viewing session 3: Season 1, Episode 3 - <i>The Shunning</i>	22
6	Week 5	Viewing session 4: Season 1, Episode 4 - <i>Success Perm</i>	22

<sup>16</sup> The week after the intervention finished, all participants took part in an eye-tracking experiment which reproduced the four conditions in the study (CF, CNF, SF, SNF) with a 7-minute video from the same TV series. Unfortunately, the analysis of the eye-tracking data falls beyond the scope of the present dissertation (see Pujadas, G. & Muñoz, C. (in prep)).



7	Week 6-7	Aptitude tests and Reading Span test [not analysed]	
8	Week 6-7	Viewing session 5: Season 1, Episode 6 - <i>Fajita Man</i>	22
9	Week 8	Viewing session 6: Season 1, Episode 7 - <i>Showdown at the Golden Saddle</i>	22
10	Week 9	Viewing session 7: Season 1, Episode 8 - <i>Phillip Goldstein</i>	22
11	Week 10	Viewing session 8: Season 1, Episode 9 - <i>License to Sell</i>	22
12	Week 11	Post-test T1	
13	Week 13	Pre-test T2	
14	Week 17	Viewing session 9: Season 1, Episode 10 - <i>Blind Spot</i>	22
15	Week 18	Viewing session 10: Season 1, Episode 11 - <i>Very Superstitious</i>	22
16	Week 19	Viewing session 11: Season 1, Episode 12 - <i>Dribbling Tiger, Bounce Pass Dragon</i>	22
17	Week 20	Viewing session 12: Season 1, Episode 13 - <i>So Chineez</i>	22
18	Week 21	Viewing session 13: Season 2, Episode 1 - <i>Family Business Trip</i>	21
19	Week 22	Viewing session 14: Season 2, Episode 2 - <i>Boy II Man</i>	21
20	Week 23	Viewing session 15: Season 2, Episode 3 - <i>Shaquille O'Neal Motors</i>	21
21	Week 24	Viewing session 16: Season 2, Episode 5 - <i>Miracle on Dead Street</i>	21
22	Week 25	Post-test T2	
23	Week 27	Pre-test T3	
24	Week 28	Viewing session 17: Season 2, Episode 7 - <i>The Big 1-2</i>	21
25	Week 29	Viewing session 18: Season 2, Episode 8 - <i>Huangsgiving</i>	21
26	Week 31	Viewing session 19: Season 2, Episode 9 - <i>We Done Son</i>	21
27	Week 32	Viewing session 20: Season 2, Episode 10 - <i>The Real Santa</i>	21
28	Week 33	Viewing session 21: Season 2, Episode 11 - <i>Year of the Rat</i>	21
29	Week 34	Viewing session 22: Season 2, Episode 12 - <i>Love and Loopholes</i>	21
30	Week 35	Viewing session 23: Season 2, Episode 13 - <i>Phil's Phaves</i>	21
31	Week 36	Viewing session 24: Season 2, Episode 14 - <i>Michael Chang Fever</i>	21
32	Week 37	Post-test T3 Questionnaire BQ3	
33	Week 37-38	Eye-tracking sessions 7 min extract from Season 2, Episode 16 - <i>Tight Two</i>	
34	Week 71-72	Delayed Post-test T3 Questionnaire BQ4	

## Chapter 5. Viewing comprehension

---

### 5.1. Introduction

Research has shown that comprehension of authentic TV programmes can be affected by a number of factors, related to the viewer, the type of on-screen text available and the characteristics of the input itself. Prior studies in the field have already observed that having access to subtitles in the learners' native language generally yields higher comprehension than having captions, as they provide on-line access to the meaning and can be read independently of the viewer's L2 skills. At the same time, it has been suggested that, for incidental vocabulary learning, captions are more beneficial than subtitles, as they expose the learners to the L2 word forms, and aid, for instance, with speech segmentation and make a first connection between the aural and written form of the words. This suggests that, at least for some aspects of language learning, captions are more beneficial than subtitles. In order for learning to take place, however, input has to be understood. Studies have often found that the benefits associated with captions are mediated by the learners' L2 proficiency and vocabulary size, suggesting that a minimum threshold might be necessary to benefit from captioning. This is especially relevant when the intended recipients of this type of input are beginner, adolescent EFL learners. While older adolescents (e.g., 15-17-year-olds) seem to be able to cope with the more demanding captions, it is still unclear whether younger, less proficient students can attain an adequate understanding of the input through L2 text.

Webb and Rodgers (2009b) pointed out that "reaching the target vocabulary size [needed for comprehension] may be too difficult a task for many learners and movies should probably not be used without providing some learning support" (Webb & Rodgers,

2009b: 420). On the basis of an analysis of the lexical coverage of different genre TV programmes, Webb (2010b) also suggested that viewers' pre-learning the most frequent low-frequency word-families in those programmes could potentially be more conducive to enhancing their comprehension than just increasing vocabulary size. Pre-teaching vocabulary that the learners will encounter in the input seems, therefore, to have a valuable educational advantage (see Chapter 6) but, to the authors' knowledge, no studies have looked at how comprehension may be affected by having additional attention demands (i.e. explicit vocabulary instruction).

Besides learners' L2 proficiency and explicit attention to vocabulary, another factor that may affect the level of comprehension of TV input is familiarity with viewing captioned and subtitled TV series in English – especially in a context such as Spain, where dubbing is still prevailing, and young learners are not generally used to viewing input in English<sup>17</sup>. It is yet to be explored, also, whether an extended exposure to TV series within the EFL classroom could have a training effect, which in turn could lead to increase the use of this media outside the formal setting. Within the classroom context, another two variables to take into consideration might be attention to and enjoyment with the TV series chosen, since – compared to viewing TV at home – students do not choose what to watch, and that could affect how input is processed (Vanderplank, 2016b). Finally, other factors that should be addressed are the level of difficulty of input itself (e.g., its vocabulary load) and the testing instruments used to assess comprehension.

---

<sup>17</sup> Although the use of online streaming platforms in Spain has increased exponentially in the last few years, at the time the present study took place, the use of these platforms was not as popular as it is nowadays. By way of example, the estimated streaming revenue of Netflix in Spain has increased from 20.6 million dollars in 2016 to 147.48 million in 2019 (source: Statista; URL: <https://www.statista.com/statistics/612043/netflix-streaming-revenues-in-spain/>).

The first study of this dissertation (i.e. Study 1) aims at addressing the aforementioned gaps and exploring the various factors that may play a role in comprehension of TV series by beginner, adolescent ELF learners. This chapter will explore TV viewing comprehension through exposure to 24 episodes over a period of eight months and, in particular, how comprehension is affected by the use of captions or subtitles, alongside variables related to the instructional focus, the learner, the lexical coverage of the episodes and the test items. It will also explore the extent to which the level of comprehension of the TV series increases as learners keep on watching more episodes.

## **5.2. Research questions**

In this chapter, the following two main research questions will be addressed:

1. To what extent does the language of the on-screen text affect comprehension of TV series? To what extent is comprehension also affected by:
  - a. Instruction-related factors (i.e. explicit focus on vocabulary items)?
  - b. Learner-related factors (i.e. general proficiency, vocabulary size, familiarity with OV, attention to and enjoyment from the series)?
  - c. Test-related factors (i.e. item format, type and difficulty)?
  - d. Episode-related factors (i.e. lexical coverage)?
  
2. Does comprehension of the episodes improve after an extended exposure to TV series (i.e. viewing 24 episodes over a period of eight months)?

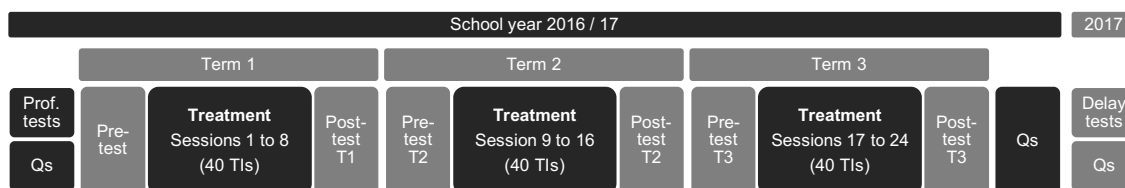
### **5.3. Participants**

All Grade 8 students took part in the one-year intervention, as it was embedded as part of a regular EFL class (see Chapter 4). From the original pool of 106 participants, however, only those who had 85% attendance or more were included in the analysis, leaving a total of 88 (56 females, 32 males). Participants in the final sample, therefore, had missed a maximum of three viewing sessions. Although episodes were self-contained, there was a general storyline throughout the season, and it was considered that missing too many episodes might have hindered their comprehension. From the four experimental groups, two had been assigned to the captions condition (n = 46) and two to the subtitles condition (n = 44). Participants who did not complete the initial proficiency tests (i.e. OPT test and X\_Lex text), and the first and third background questionnaire (i.e. BQ1 and BQ3) had to be excluded from part of the analysis.

### **5.4. Materials and procedure**

A complete description of the intervention materials can be found in Chapter 4. The instruments used for Study 1 are highlighted in a darker colour in Figure 5.1 below. They were, namely, the 24 comprehension tests completed after each episode (section 4.4.2.3), the two measures of language proficiency collected at the beginning of the school year (section 4.4.1), and two background questionnaires – one administered before the beginning of the intervention (section 4.4.3.1) and one at the end (see section 4.4.3.2), which were used to assess the level of familiarity, attention and enjoyment.

Figure 5.1. Materials used in Study 1



Comprehension tests had a total of 10 items, combining two formats – five multiple-choice items (MC) and five true/false items (TF). They also included two types of information: textually explicit items (TE) and inferential items (IN). The OPT test was used as a general measure of proficiency, and the X\_Lex test was used as a measure of vocabulary size. Familiarity with viewing captioned and subtitled audio-visual input in English was assessed prior to the intervention through BQ1#2. Attention to and enjoyment from the series was assessed at the end of the intervention through BQ3#5 and BQ3#6.

## 5.5. Preliminary analysis

### 5.5.1. Learner-related variables

Table 5.1 reports the descriptive statistics for initial proficiency (OPT scores), vocabulary size (X\_Lex scores)<sup>18</sup> and familiarity with viewing OV input (prior to the intervention). It reports the mean scores per each language group, with standard

---

<sup>18</sup> The RANGE software (used to analyse the lexical coverage of the input) and the X\_Lex test (used as the measure for learners' vocabulary size) are based on different word-lists, but a validation study by Miralpeix (2012) has shown that the results of the Levels Test and the X\_Lex are comparable. Also, while the RANGE software indicates the coverage by word-level, the X\_Lex test provides a total score of out 5,000 words (by adding up the knowledge in each of the 5 word-bands). However, it seems logical to assume that – out of a score of 2,000 words – most of the words would be from the first or second thousand word-bands, even if some of the words do indeed belong to the fourth or fifth bands (Miralpeix, personal communication, December 18, 2018). An analysis of the X\_Lex results in each word band in the present study also revealed this tendency (75% of the words were from the first three bands).

deviations (SD) shown in brackets. As can be observed, there were no significant differences between the groups in terms of proficiency ( $F(1,77) = .861$ ,  $p = .356$ ), vocabulary size ( $F(1,74) = .203$ ;  $p = .653$ ) or familiarity with viewing OV ( $F(1,83) = .015$ ;  $p = .904$ ).

Table 5.1. Initial proficiency, vocabulary size and familiarity with OV per language condition

	n	Proficiency (OPT)	n	Vocabulary size (X_Lex)	n	Familiarity with OV <sup>a</sup>
<b>Subtitles</b>	37	92.76 (13.85)	38	1,988 (486)	41	2.02 (.79)
<b>Captions</b>	42	95.79 (15.00)	38	1,931 (601)	44	2.05 (.81)
<b>Mean</b>		94.37 (14.47)		1,959 (544)		2.04 (.79)

<sup>a</sup> Familiarity with viewing Original Versions in English was re-categorized in three levels of similar size: low-frequent viewers (1), mid-frequent viewers (2), and high-frequent viewers (3) (see below).

The measure for general proficiency was re-categorized into three levels according to the OPT scores, distributing participants in three CEFR groups: Pre-A, A1 and A2/B1. Familiarity with OV input was calculated by combining the reported frequency of viewing audio-visual input in English with or without on-screen text – assessed at the beginning of the school year. For the analysis, the original responses were recoded into three general categories – with a similar number of participants in each group – according to the approximate weekly exposure to English audio-visual input: low-frequent viewers (1 to 2 hours per week), mid-frequent viewers (from 2 hours per week to 1 hour-daily exposure), and high-frequent viewers (over an hour of daily exposure). The variables of attention to and enjoyment with the TV series – originally rated in a 5-point scale – were also recoded into three levels: low (which collapsed “not at all”, “not much” and “a bit” categories), mid (“quite a lot”) and high (“a lot”). This was done to have a more balanced number of learners in each group for the analysis. Figure 5.2 shows the distribution of participants in the recoded categories.

Table 5.2. Distribution of participants for the categorical variables

Factors	Levels	Subtitles		Captions	
		n	%	n	%
Proficiency	Pre-A	14	37.8	13	31
	A1	16	43.2	20	47.6
	A2 / B1	7	18.9	9	21.4
Familiarity	Low	12	29.3	13	29.5
	Medium	16	39	16	36.4
	High	13	31.7	15	34.1
Attention	Low	1	2.4	7	15.6
	Medium	14	34.1	28	62.2
	High	26	63.4	10	22.2
Enjoyment	Low	4	9.8	23	51.1
	Medium	18	43.9	15	33.3
	High	19	46.3	7	15.6

### 5.5.2. Episodes' lexical coverage

The analysis of the scripts of the 24 episodes showed that, as mentioned before, the series reached a lexical coverage of 95.70% overall – which is the general threshold for adequate comprehension (e.g., Van Zeeland & Schmitt, 2013) – at the 3k word-level, plus marginal words and proper nouns (see section 4.3.3). Exploration of the data showed, however, that coverage provided by the most frequent 3,000 word families of the BNC/COCA word lists ranged from 93.98% to 96.73% between the episodes. Although the variance seems small (the largest difference was 2.75% between the episodes with the lowest and highest coverage), research has suggested that even a small increase in lexical coverage can benefit comprehension (Laufer & Ravenhorst-Kalovski, 2010). Therefore, the percentage of lexical coverage per episode was also included as an episode-related factor in the analysis.



### 5.5.3. Scoring and validation of the comprehension tests

All 240 comprehension items from the 24 tests were scored dichotomously (1 = correct / 0 = incorrect). Once tests scores were obtained, the difficulty index was calculated to assess how easy or hard the items were. This index is calculated in relation to the total correct responses of an item within the sample, and the higher the index the easier the item is – the more correct responses has it received (Del Rincón, Arnal, Latorre & Sans, 1995). Table 5.3 shows the number of items in each index level and the percentage they represent within the 250 items. As can be observed, around 40% were very easy, 42% had an easy to medium level of difficulty and 18% were considered hard or very hard.

Table 5.3. Distribution of items by difficulty index

Difficulty index	N items	%
Very Easy (+0.75)	95	39.58
Easy (0.74 to 0.55)	75	31.25
Medium (0.54 to 0.45)	28	11.67
Hard (0.44 to 0.25)	31	12.92
Very Hard (-0.25)	11	4.58
<b>Total</b>	<b>240</b>	<b>100</b>

An item discrimination index was used as a validation measure to calculate the extent to which success in an item corresponded to success on a whole test (Kelly, 1939). This index reports the difference between the proportions of high and low test-takers answering correctly a dichotomous item. The higher the index, the better the item. This index showed that – out of 240 items – 67.1% were very good, 12.5% were good, 11.3% were regular and 9.2% were poor discriminators (see Table 5.4). Items that were poor discriminators were not eliminated upon checking that they were homogeneously

distributed across the 24 tests. The mean discrimination index per test was “good” for 2 tests and “very good” for the other 22 tests.

*Table 5.4. Distribution of items by discrimination index*

<b>Discrimination index</b>	<b>N items</b>	<b>%</b>
Very Good (+0.40)	161	67.08
Good (0.30 to 0.39)	30	12.50
Regular (0.20 to 0.29)	27	11.25
Poor (-0.19)	22	9.17
<b>Total</b>	<b>240</b>	<b>100</b>

## 5.6. Results

### 5.6.1. Overall comprehension with the support of captions and subtitles

Table 5.5 shows the overall mean comprehension percentage for each language condition – the mean percentage of comprehension across the 24 episodes viewed. Learners in the subtitles group obtained around 82% comprehension, while learners in the captions group achieved a 64% comprehension rate. It is interesting to note that there was a wide gap (around a 50% difference) between the most and least successful participants in both language conditions.

*Table 5.5. Mean comprehension scores by language condition*

	<b>n</b>	<b>Mean (%)</b>	<b>SD</b>	<b>Minimum</b>	<b>Maximum</b>
<b>Subtitles</b>	42	81.54	9.09	50.91	93.48
<b>Captions</b>	46	64.18	11.14	43.48	90.45
<b>All</b>	88	72.47	13.38	43.48	93.48

### 5.6.2. Factors affecting TV Viewing Comprehension

The first research question aimed at examining the effect of several variables on comprehension scores, related to the intervention, the learners, the testing instruments and the episodes selected for the intervention. Table 5.6 shows the 10 factors, including the different levels found in the categorical variables.

*Table 5.6. Independent factors*

Intervention	1. Language of the on-screen text: captions   subtitles 2. Type of instruction: focused   non-focused
Learner	3. Proficiency: Pre-A   A1   A2/B1 4. Vocabulary size 5. Familiarity with OV: low-frequent   mid-frequent   high-frequent viewers 6. Attention: low   medium   high 7. Enjoyment: low   medium   high
Test	8. Comprehension items' format: multiple-choice   true-false 9. Comprehension items' type: textually explicit   inferential
Episode	10. Lexical coverage of the episodes

A Generalized Linear Mixed Model (GLMM) with repeated measures was calculated using SPSS 21.0 with comprehension score as the outcome variable, and the 10 variables abovementioned as fixed factors, including all two-way interactions. This type of statistical test was found particularly appropriate because of several reasons: it does not require normal distribution nor homogeneity of variances; there was an acceptable ratio of observations to independent variables; and there was no multicollinearity. GLMM also allows the inclusion of learner variables, intervention variables and item variables in a single model. In this type of analysis, a particular score (correct or incorrect) is defined

by the combination of participant, item and response. The GLMM was based on 17,310 observations.

Table 5.7 shows the number (and the percentage in brackets) of correct and incorrect responses for the 240 comprehension items, separated by language condition. As can be observed, the subtitles group had overall 17.8% more correct responses than the captions group.

*Table 5.7. Number of correct and incorrect responses per language condition*

	<b>Correct responses</b>	<b>Incorrect responses</b>
<b>Subtitles</b>	7.834 (82.0%)	1.716 (18.0%)
<b>Captions</b>	6.768 (64.2%)	3.772 (35.8%)
<b>Total</b>	14.602 (72.7%)	5.488 (27.3%)

To arrive at the best fitting model, all the explanatory variables were entered in the model. All non-significant interactions and main effects ( $p < .10$ ) were then removed one by one. Table 5.8 presents the final fitted model, and Table 5.9 shows the significant main effects for the categorical variables. The model revealed that there were four factors that significantly contributed to the model ( $p < .05$ ): language of the on-screen text, vocabulary size, item format and lexical coverage; and one factor that contributed marginally – type of instruction. Three significant interactions emerged: between language and vocabulary size, between language and item format and between language and item type.

Table 5.8. Results from the GLMM

Terms	Coeff	SD	t	Sig	Exp Coeff	95% CI for Exp Coeff	
						Lower	Upper
Language <sup>a</sup>	1.610	.4512	3.569	<.001	5.004	2.067	12.116
Instruction <sup>b</sup>	-.189	.1107	-1.708	.088	.828	.666	1.028
VS	.001	.0001	4.499	<.001	1.001	1.000	1.001
Format <sup>c</sup>	-.187	.0447	-4.187	<.001	.829	.760	.905
Coverage	.173	.0241	7.162	<.001	1.188	1.134	1.246
Lang*VS	.000	.0002	-2.025	.043	1.000	.999	1.000
Lang*Format	.184	.0715	2.570	.010	1.202	1.045	1.383
Lang*Type	.279	.0716	3.900	<.001	1.322	1.149	1.522

Language = Language of the on-screen text; Instruction = Type of instruction; VS = vocabulary size; Format = Item format; Type = Item type; Difficulty = Difficulty index; Exp Coeff = Exponential Coefficient; CI = Confidence Interval

<sup>a</sup> Reference category: subtitles

<sup>b</sup> Reference category: focused instruction

<sup>c</sup> Reference category: MC item type

Table 5.9. Results from the GLMM: influence of fixed main effects

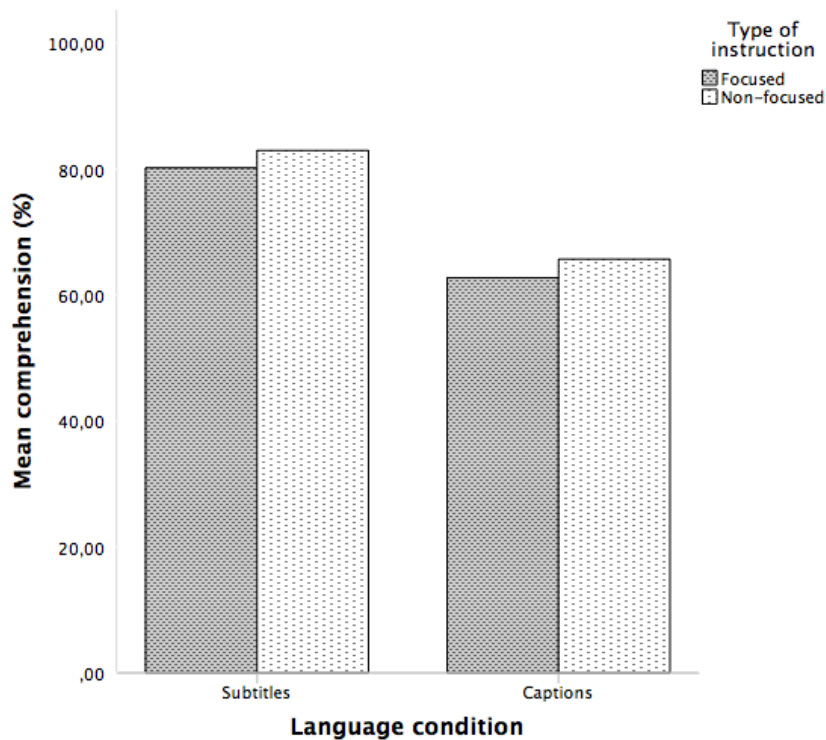
Factor	Group	$\bar{X}$ (SE)	df	F	p value
Language	Subtitles	.826 (.013)	1, 17300	76.355	<.001
	Captions	.646 (.016)			
Instruction	Yes	.728 (.015)	1, 17300	2.923	.087
	No	.764 (.015)			
Format	MC	.737 (.011)	1, 17300	7.081	.008
	TF	.755 (.011)			

MC = multiple-choice; TF = true-false; TE = textually explicit; IN = inferential

### 5.6.2.1. Intervention-related variables

The GLMM analysis showed that there was a significant main effect of language of the on-screen text ( $p < .001$ ), indicating that an average learner's score would be 14% higher if they had access to L1 subtitles when all other factors were held constant (see Table 5). There was an interaction effect between language condition and three other parameters – vocabulary size, item format and item type –, which suggested that the effect of these three variables needs to be explained in relation to language (see below). There was a small tendency for comprehension scores to depend on type of instruction, indicating that the two groups who received explicit instruction on target vocabulary items tended to score lower than the two groups without instruction, although differences between groups did not reach statistical significance ( $p = .088$ ). Figure 5.2 illustrates the effect of both variables.

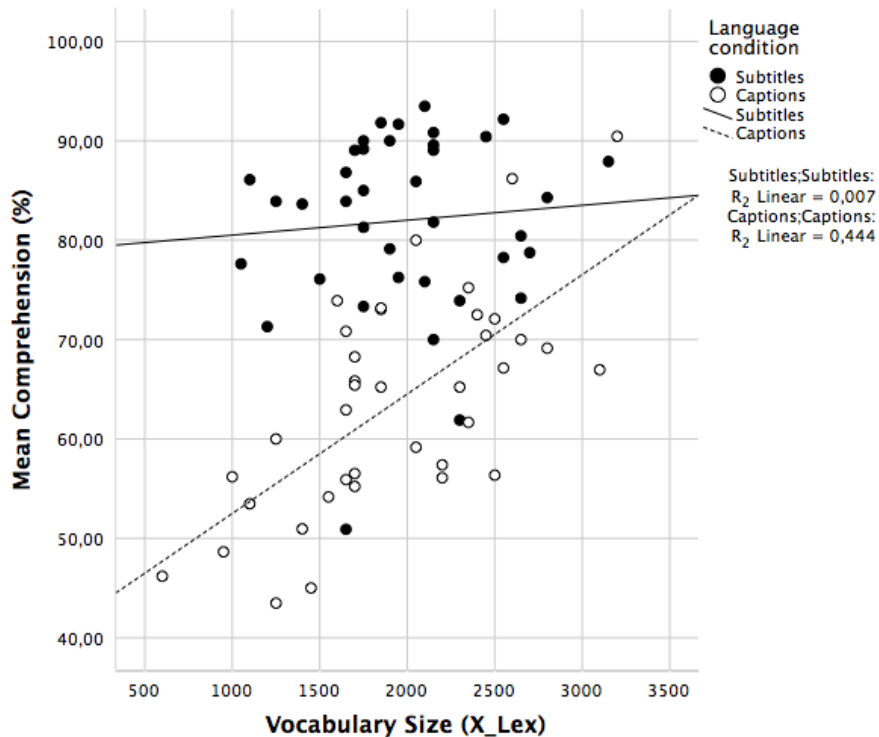
Figure 5.2. Mean comprehension by language and instruction condition



### 5.6.2.2. Learner-related variables

The GLMM model showed that vocabulary size was the only learner-related variable in our study that emerged as a predictor of comprehension ( $p = .004$ ), while general proficiency did not appear to be a significant predictor (although the two variables correlate significantly;  $r = .334$   $p < .001$ ). The interaction effect found between language group and vocabulary size ( $p = .043$ ) indicated that the effect of vocabulary size depended on the language of the on-screen text, and that vocabulary size was only significant for the captions group. Figure 5.3 illustrates the relationship between these two variables, showing the participants' average percentage of comprehension per vocabulary size and language condition.

Figure 5.3. Mean comprehension by vocabulary size



As can be observed, there were few participants in the captions conditions who scored over 80%, which was the mean percentage of comprehension attained by participants in the subtitles group. It can also be noticed that – in both conditions – there were several participants with a vocabulary size over 2,500 words who scored below classmates with smaller vocabularies.

Attention and enjoyment did not appear to contribute to the final fitted model, but it was observed that, when introduced in the model individually, they emerged as significant predictors ( $p < .001$  and  $p = .003$  respectively), with higher attention and enjoyment associated with higher comprehension gains. Although no interaction was found with language condition, further exploration showed that attention and enjoyment were significantly higher in the subtitles group compared to the captions group ( $F(3,82) = 6,581; p < .001$  and  $F(3,82) = 8,753; p < .000$ , respectively). Familiarity with viewing OV did not contribute to explaining comprehension scores.

#### 5.6.2.3. *Test-related variables*

With respect to test-related variables, the GLMM indicated that item format was a strong predictor of comprehension, while item type was not (see Table 5.8 above). Both variables, however, interacted significantly with language condition, which indicates that their effect has to be interpreted in relation to the language of the on-screen text. Table 5.10 presents the mean comprehension scores per item format and item type when divided by language condition.



Table 5.10. Results from the GLMM: interactions of categorical factors

Interaction			$\bar{X}$ (SE)	df	F	p value
Language* Format	Subtitles	MC	.826 (.013)	1, 17300	.004	.952
		TF	.826 (.013)			
	Captions	MC	.624 (.017)			
		TF	.667 (.017)			
Language* Type	Subtitles	TE	.835 (.013)	1, 17300	5.125	.024
		IN	.817 (.014)			
	Captions	TE	.628 (.017)			
		IN	.663 (.017)			

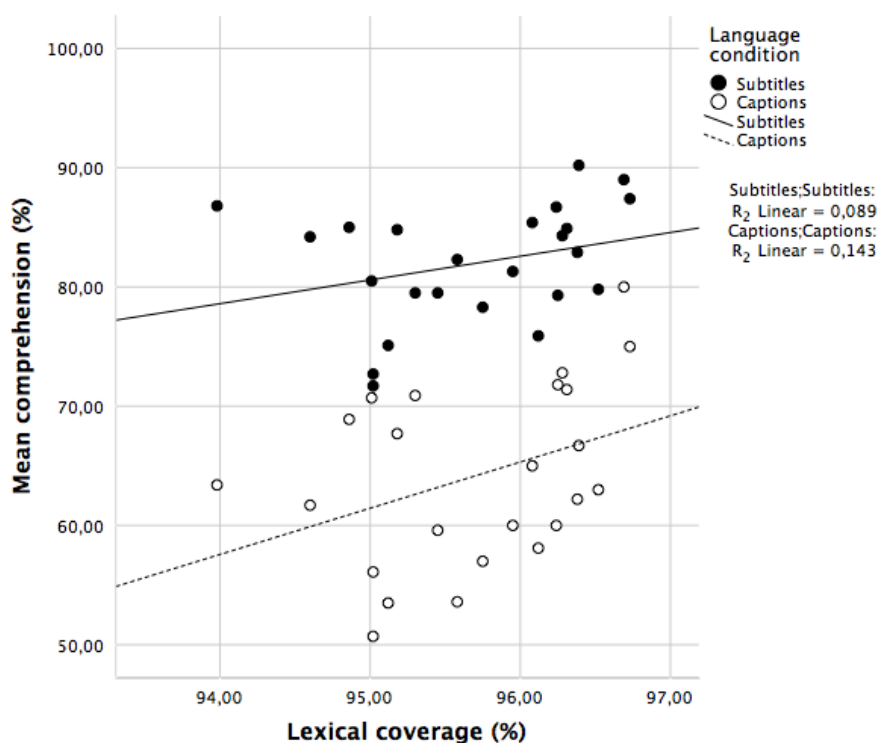
MC = multiple-choice; TF = true-false; TE = textually explicit; IN = inferential

Regarding item format, participants had more corrected responses in the true-false items (75.5%) than in the multiple-choice items (73.7%), a small but significant difference ( $p < .001$ ). When language condition was taken into account, however, it was observed that learners in the subtitles groups performed equally in both formats (82.6%), while in the captions groups learners performed significantly better in the true-false items (66.7%) than the multiple-choice items (62.4%) ( $p < .001$ ). On the other hand, there was no main effect of item type, but the interaction with language condition indicated that correct responses in one type or the other depended on the language of the on-screen text. The subtitles condition performed better in the textually explicit items (83.5%) than in the inferential items (81.7%) ( $p = .024$ ). Inversely, the captions group performed better in the inferential items (62.8%) than the textually explicit ones (66.3%) ( $p = .001$ ).

#### 5.6.2.4. Episode-related variables

The lexical coverage of the episodes also emerged as a strong predictor of comprehension ( $p < .001$ ), independently of the language the on-screen text, with the items in episodes with higher percentage of lexical coverage receiving more correct responses. Figure 5.4 illustrates the positive linear relationship between the percentage of lexical coverage and comprehension by language condition (in the scatterplot, the dots represent the mean percentage of comprehension per each of the 24 episodes). Although no interaction with language condition emerged in the model, it can be observed that the lexical coverage tends to have a slight stronger effect when captions are available.

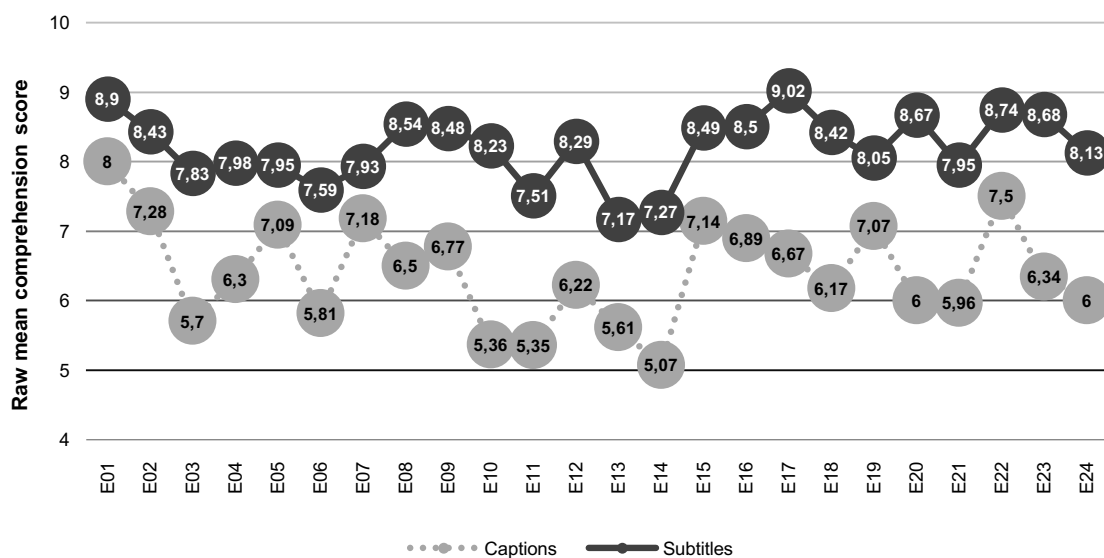
Figure 5.4. Episodes' mean comprehension by episodes' lexical coverage



### 5.6.3. Changes in comprehension over time

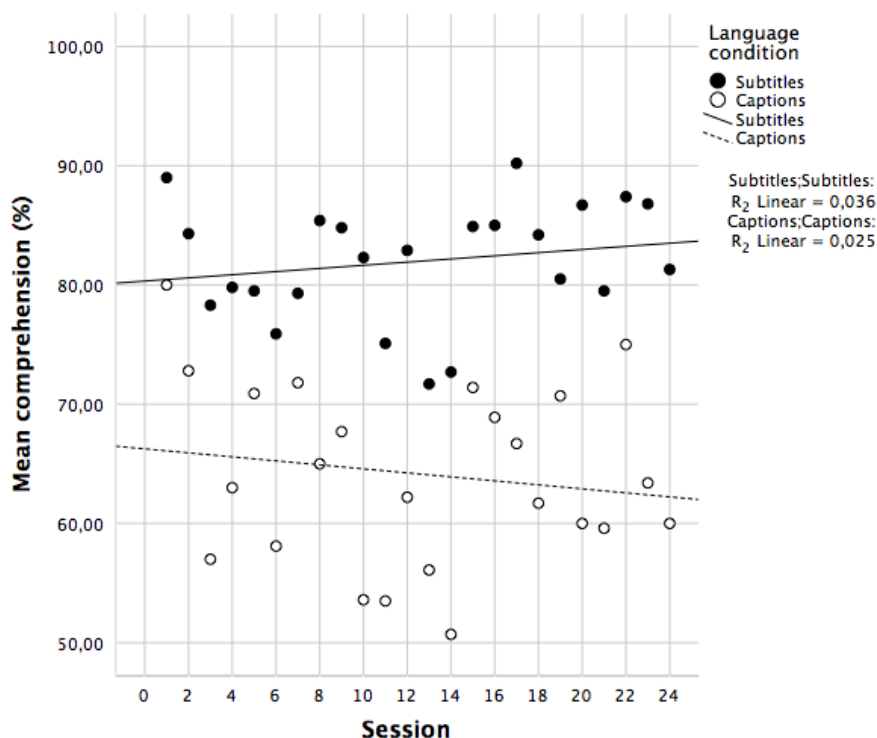
The second research question sought to explore if comprehension had improved over the course of the study. Figure 5.5 shows the mean raw score (out of 10) for the 24 episodes by each language group (captions and subtitles). Independent samples t-tests revealed that the subtitles group significantly outperformed the captions group in all 24 episodes – ranging from  $p < .001$  to  $p = .021$ . While comprehension scores vary from episode to episode, there appears to be no linear progression from the first to the last viewing session.

Figure 5.5. Mean comprehension throughout the 24 episodes



To assess changes over time, a linear regression was run with episode comprehension as the criterion variable, and viewing session (from 1 to 24) as the independent factor. Results showed that session did not predict comprehension ( $\beta = .006$ ,  $p < .939$ ). The same lack of predictability was found when the analysis was run for both language groups separately, as Figure 5.6 below illustrates. In the figure, the dots represent, again, the mean comprehension of each episode per each language condition.

Figure 5.6. Episodes' mean comprehension by session



A visual exploration of the data in Figure 5.5 suggests, nevertheless, that the two lines follow a similar pattern – with the exception of few episodes. Data was further explored at episode level, by looking at the episodes with the lowest and highest percentage of comprehension (the 25% percentiles, that is, the six worst and the six best episodes). Table 5.11 below shows the mean raw comprehension scores (out of 10) per each of the 24 episodes divided by language group, as well as information on each episode's lexical coverage. The “▼” and “▲” symbols indicate, respectively, the six episodes with lowest and highest values for lexical coverage and comprehension.

It was observed that five out of six episodes with the lowest percentage of comprehension in both languages were the same (E03, E06, E11, E13, E14), and two of them (E13 and E14) were amongst the episodes with lowest coverage. On the other hand, when looking at the top six episodes, only two out of six were the same (E01 and E22), but both were the ones with the highest lexical coverage. Additionally, it could also

be observed that, although lower percentages of lexical coverage were found in episodes viewed in T2 and T3, the lowest mean scores in comprehension were found in T1 and T2, while the highest mean scores in comprehension can be generally found by T2 and T3 – with the exception of the two first episodes (i.e. E01 and E02).

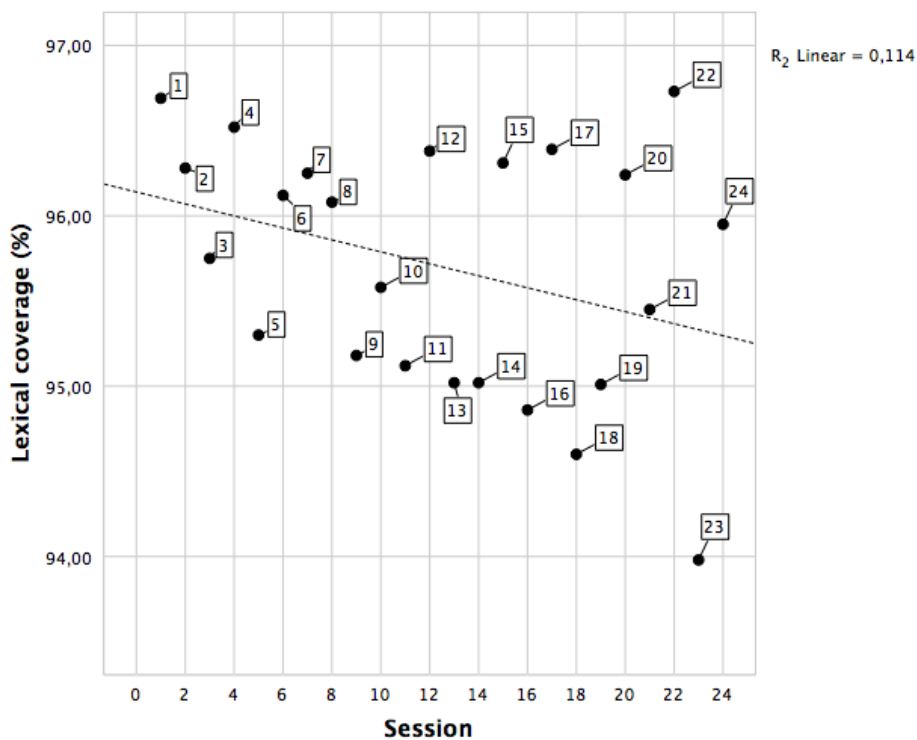
Table 5.11. Comprehension scores per episode and language condition

Term	Episode (session)	Lexical coverage	Mean raw comprehension scores (out of 10)		
			Subtitles	Captions	Mean
T1	1	96.69 ▲	8.9 ▲	8 ▲	8.45 ▲
	2	96.28	8.43	7.28 ▲	7.86 ▲
	3	95.75	7.83 ▼	5.7 ▼	6.77 ▼
	4	96.52 ▲	7.98	6.3	7.14
	5	95.3	7.95	7.09 ▲	7.52
	6	96.12	7.59 ▼	5.81 ▼	6.7 ▼
	7	96.25	7.93 ▼	7.18 ▲	7.56
	8	96.08	8.54 ▲	6.5	7.52
T2	9	95.18	8.48	6.77	7.63
	10	95.58	8.23	5.36 ▼	6.8 ▼
	11	95.12	7.51 ▼	5.35 ▼	6.43 ▼
	12	96.38 ▲	8.29	6.22	7.25
	13	95.02 ▼	7.17 ▼	5.61 ▼	6.39 ▼
	14	95.02 ▼	7.27 ▼	5.07 ▼	6.17 ▼
	15	96.31 ▲	8.49	7.14 ▲	7.81 ▲
	16	94.86 ▼	8.5	6.89	7.7 ▲
T3	17	96.39 ▲	9.02	6.67	7.85 ▲
	18	94.6 ▼	8.42	6.17	7.3
	19	95.01 ▼	8.05	7.07	7.56
	20	96.24	8.67 ▲	6	7.34
	21	95.45	7.95	5.96	6.96
	22	96.73 ▲	8.74 ▲	7.5 ▲	8.12 ▲
	23	93.98 ▼	8.68 ▲	6.34	7.51
	24	95.95	8.13	6	7.07
<b>Mean</b>		<b>95.70</b>	<b>8.20</b>	<b>6.42</b>	<b>7.31</b>

Differences between the academic terms was also explored. The mean percentage of comprehension was 74.19% in T1 (N = 87; SD = 12.15), 68.92% in T2 (N = 81; SD = 14.92), and 74.08% in T3 (N = 81; SD: 14.92). A related samples Friedman's 2-way analysis of variance by ranks revealed that there was a statistical significant differences in comprehension across the three terms ( $X^2(2) = 32.276, p < .001$ ). Post-hoc tests showed, however, that differences were only significant from T1 to T2 ( $Z = 4.759, p < .001$ ), and from T2 to T3 ( $Z = -4.924, p < .001$ ), but not between T1 and T3 ( $Z = -.166, p = 1$ ). When the analysis was run per each language condition separately, results were the same: there were no significant difference between T1 and T3 comprehension percentage in either the subtitles ( $Z = -1.945, p = .155$ ) nor the captions group ( $Z = 1.686, p = .276$ ).

Episodes with a lower coverage seemed to be condensed in the second half of the intervention (i.e. from episode 13 onwards). In view of this, a last analysis was done to explore the relationship between lexical coverage and the order of the sessions (from 1 to 24). A regression analysis showed that session did not significantly predict lexical coverage ( $F(1,22) = 2.831, p = .107, R^2 = .114$ ), although there seemed to be a tendency for lexical coverage to be lower towards the end of the intervention, as illustrated in Figure 5.7 below.

Figure 5.7. Episodes' lexical coverage and session



#### 5.6.4. Comprehension of episodes

Table 5.12 provides more detailed information on the mean comprehension score (out of 10) for the 24 episodes, including information regarding the minimum and maximum scores reached by learners in each language condition.

Table 5.12. Comprehension scores by episode and language condition

Episode	Subtitles					Captions				
	N	Mean	SD	Min	Max	N	Mean	SD	Min	Max
1	41	8.9	1.48	3	10	45	8	1.33	5	10
2	40	8.43	1.55	3	10	46	7.28	1.41	2	10
3	40	7.83	1.48	3	10	46	5.7	1.46	3	9
4	42	7.98	1.73	2	10	46	6.3	1.72	2	9
5	40	7.95	1.30	3	10	44	7.09	1.54	4	10
6	41	7.59	1.73	3	10	43	5.81	1.71	3	10
7	41	7.93	1.40	4	10	45	7.18	1.53	4	10
8	39	8.54	1.30	4	10	44	6.5	2.07	1	10
9	40	8.48	1.50	4	10	43	6.77	1.69	3	9
10	40	8.23	1.44	3	10	42	5.36	2.32	1	9
11	41	7.51	1.79	3	10	43	5.35	1.80	2	9
12	42	8.29	1.60	4	10	45	6.22	2.12	3	10
13	36	7.17	1.70	0	9	44	5.61	1.99	1	9
14	37	7.27	1.45	4	9	45	5.07	1.71	2	10
15	39	8.49	1.17	5	10	44	7.14	1.76	2	10
16	40	8.5	1.09	6	10	44	6.89	1.81	2	10
17	41	9.02	0.82	7	10	45	6.67	1.76	3	10
18	38	8.42	1.37	5	10	42	6.17	1.71	3	9
19	39	8.05	1.91	1	10	44	7.07	1.52	4	10
20	39	8.67	1.58	4	10	41	6	2.63	0	10
21	38	7.95	1.71	4	10	46	5.96	1.81	3	10
22	42	8.74	1.19	5	10	44	7.5	1.66	4	10
23	40	8.68	1.51	4	10	41	6.34	2.13	2	10
24	39	8.13	1.44	3	10	42	6	2.07	2	10

As can be observed looking at the maximum raw scores per episode, there were learners in both language conditions who scored a 100% in the comprehension tests (10 out of 10). In the subtitles groups, from the 955 tests completed there were 185 tests (19.37%) with maximum score. In contrast, in the captions groups, only 37 out of 1,054 test had a perfect score (3.51%). Around 50% of the tests with best scores were attained in T3 for both language groups.



A closer look at the most successful students in the captions condition revealed that, unsurprisingly, they were the ones who obtained a 100% comprehension in several episodes. The most successful student was also the one with the highest comprehension score in the captions condition and the one with the highest initial vocabulary size within the entire sample (i.e., 3,200 words). Table 5.13 provides details of the top-five learners' performance in the captions group.

*Table 5.13. Top five higher-achievers in the captions condition*

ID	Test with 100% score	Mean % Comprehension	Ranking in captions group	Ranking in the sample	Proficiency level (OPT test)	Vocabulary size (X_Lex)	Explicit focus on vocabulary
312	8 / 24	90.45%	1	6	B1	3,200	No
112	4 / 24	86.19%	2	16	Pre-A	1,850	Yes
104	4 / 24	84.17%	3	22	A1	-	Yes
114	4 / 24	81.25%	4	28	A2	-	Yes
317	2 / 24	75.24%	5	31	-	2,050	No

## 5.7. Summary of findings

The main findings from Study 1 can be summarized as follows:

1. Learners with access to subtitles had significantly higher comprehension scores (17.8% higher) than learners with access to captions.
2. Explicit instruction on target lexical items did not significantly affect comprehension, but there was a tendency for the focused groups to have lower scores.
3. Vocabulary size was found to be a strong predictor of comprehension, with larger vocabularies leading to higher comprehension. This was only found, however, when the episodes were viewed with captions.

4. True/false items were found to be easier than multiple-choice items, although this was only significant in the captions condition.
5. The type of information was processed differently according to the language available in the on-screen text; learners with access to subtitles scored better in textually explicit items, while learners with access to captions did better at inferential items.
6. Lexical coverage was also found to be a strong predictor of comprehension, with higher lexical coverage related to higher comprehension.
7. No pattern of improvement could be observed from the first to the last episode.

## **5.8. Discussion**

### **5.8.1. Language and instruction**

Results showed that language of the on-screen text was a significant predictor of comprehension scores, with the subtitles group significantly outperforming the captions group, as expected. This was found consistently across the 24 episodes of the TV series (*cf.* Rodgers, 2013; Rodgers & Webb, 2017), indicating that, regardless of other mediating factors such as the lexical coverage of the episodes, viewing TV series with access to the L1 yields higher comprehension than viewing them with captions. Previous studies have also showed the advantages of having L1 text for comprehension in audio-visual media (Bianchi & Ciabattini, 2008; Birulés-Muntané & Soto-Faraco, 2016; Latifi et al., 2011; Lwo & Lin, 2012; Markham et al., 2001; Markham & Peter, 2003).

Participants in the subtitles condition attained 82% mean comprehension. Although it would seem that having access to the native language would lead to 100% understanding of the dialogue, research shows that it is uncommon. In studies comparing

the use of captions and subtitles – at different levels of proficiency –, the mean comprehension for the subtitles groups were 67% (Markham & Peter (2003), 72% (Bianchi & Ciabattini, 2008), 72% (Latifi, et al., 2011), 82% (Markham et al., 2001), and 93% (Birulés-Muntané & Soto-Faraco, 2016). There may be several reasons for this (e.g., factors related to the level of detail of the questions, working memory), but that falls beyond the scope of the study. In the current study, there were participants who obtained a 100% score on several sessions, and some were viewing the episodes with captions. Although participants in the captions condition had a significantly lower comprehension, they nevertheless reached a considerable level of understanding of the content (i.e. 64%). Setting aside the differences in types of input and testing instruments, the percentage of comprehension for the captions group was similar to that found in other studies with adult learners. This is an important finding, since it would confirm that learners at this age and proficiency could already benefit from captions.

Results also showed that having explicit instruction on target vocabulary had a small negative effect on overall comprehension – a drawback also found in previous studies with advanced organizers (e.g., Lee, 2007) –, indicating that learners at this age and proficiency level may find it hard to split their attention between the two demands (VanPatten, 1996, 2002). This suggests that research assessing comprehension performance when participants are also asked to pay attention to language forms (e.g., vocabulary, grammar) might need to take into account the depleting effects that explicit attention to specific aspects of the language might have on students' performance. Yet, it has also been found that – in this context – directing learners' attention towards target vocabulary renders significant improvement in vocabulary recall (see Chapter 6). In the current study, the effect of instruction did not significantly hinder comprehension, so even if comprehension might have been slightly lower, it seems safe to say that pre-teaching vocabulary is beneficial for language learning. Possible trade-offs between content

comprehension and learning specific language aspects, such as vocabulary, deserve further attention and exploration.

### **5.8.2. Factors related to the learner**

Although general proficiency did not emerge as a predictor in our study, it was found that vocabulary size was positively related to comprehension scores, with larger vocabulary related to higher comprehension. This falls in line with results from other studies that also found that learners' vocabulary knowledge was a good predictor of comprehension (e.g., Montero-Perez, Peters, Clarebout & Desmet, 2014; Vulchanova et al., 2015) – which concurs with findings on vocabulary acquisition research (e.g., Peters & Webb, 2018). The interaction between vocabulary size and language indicated, however, that vocabulary size was a significant predictor only in the captions condition. This may suggest that learners in the subtitles condition relied more on reading the L1 text than on listening to the L2 audio (Steward & Pertusa, 2004; Vandergrift, 2007), thus making prior L2 lexical knowledge less relevant for comprehension when having subtitles available. While this finding points out the value of subtitles as a scaffold for lower-proficiency learners for accessing content, it also implies that learners with access to the L1 might have not been processing the L2.

Our results partially concur with those of Lwo and Lin (2012), who found that Grade 8 learners – of same age as participants in the current study – benefitted better from subtitles than captions. In their study, however, Lwo and Lin acknowledged that the subtitles group was more proficient, a setback not found in the present study, in which both groups were comparable in terms of initial proficiency and vocabulary size. On the other hand, these results contrast with results from two previous studies with adolescent learners, in which it was found that there were no significant differences between

language conditions (Vulchanova et al., 2015), and that the on-screen text had a distracting effect for more advanced students (Baristow & Lavour, 2011). Yet, participants in those studies were older and, probably, more proficient. It is possible that, with an increase in proficiency and vocabulary size, the difference between our language groups would have been smaller.

The other three learner-related factors – familiarity with viewing OV, attention to, and enjoyment from the TV series – did not appear to predict comprehension outcomes. A reason might be that the effect of these variables might have been overpowered by other factors (e.g., language condition, lexical coverage). It is also likely that participants in the current sample did not have as much prior experience viewing OV as for this factor to have a significant effect on comprehension scores. Although attention and enjoyment did not emerge as significant predictors, both were significantly higher in the subtitles groups. It might have been the case that language condition overpowered these other two parameters in the analysis, or it might be that higher attention to and enjoyment from the TV series were a result of the language condition – they were higher *because* learners were viewing the programmes with access to the L1 text (see Chapter 8).

### **5.8.3. Factors related to testing instruments and the audio-visual input**

Regarding test characteristics, item format was revealed to predict comprehension scores, with TF items having more correct responses than MC items. Yet, once language of the on-screen text was taken into account, the difference was only significant for the captions groups, suggesting that the availability of the L1 rendered item format unimportant. The language of the on-screen text also mediated responses by item type. While overall comprehension scores were not affected by item type, once language of the on-screen text was taken into account, for the subtitles groups it was found that

recalling textually explicit information was easier than recalling inferential information. This falls in line with findings in the listening research literature showing that processing scattered information is harder than recalling information from just one location, and that recalling exact content tends to be easier than recalling the gist or main idea (Buck, 2001). However, for the captions group it was the inferential items that received significantly more correct responses. It could be the case that – for the captions group – answering textually explicit items demanded that learners understood details that they might have missed due to the fast speech rate of the series and their low L2 linguistic skills (i.e., they could not use bottom-up processing successfully), whereas for inferential items the fact that they could gather information from different parts compensated for a missed piece of data (i.e., they were more successful at using top-down processing).

Another parameter that had a significant effect on comprehension was the episodes' lexical coverage, which has been shown to be a strong predictor in past research in listening comprehension (e.g., Hu & Nation, 2000) and video comprehension (e.g., Rodgers, 2013). Even if the difference between episodes was relatively small, episodes with higher lexical coverage had more percentage of correct responses. While the complexity of the plot or the familiarity with the topic of individual episodes (which sometimes included culture-bound references such as Thanksgiving) might have also played a role, episode lexical coverage appears to be a reliable and robust predictor for comprehension, independently of other factors such as on-screen text language or learners' vocabulary size.

#### **5.8.4. Changes in comprehension over time**

Comprehension rates across the whole eight-months intervention was analysed from different perspectives, to assess whether the level of comprehension of the episodes

increased with successive viewings, as suggested in the literature (Rodgers, 2013; Vanderplank, 2019; Webb, 2011). No clear pattern of improvement could be observed, however, as more episodes were viewed. Time (i.e. viewing session) did not predict comprehension, suggesting that comprehension was episode-dependent. Instead, lexical coverage seemed to be the best predictor in explaining the differences in comprehension. The fact that the episodes with the lowest comprehension were the same across language condition might also point in this direction.

An analysis of the distribution of the episodes with lowest and highest percentage of lexical coverage across the study also revealed that the episodes with lower lexical coverage (and therefore with higher vocabulary load) were found mostly in the second half of the intervention, which could explain why no improvement in comprehension rates could be observed. While learners might have been getting better at processing and understanding the TV series, the episodes they were watching were increasingly slightly more difficult. This would be supported, additionally, by learners' perceptions at the end of the intervention, as 74% of students reported that they understood better the series by then (see Chapter 8). It could also be the case, then, that their comprehension was starting to improve, but it was not yet detected by the measures used in the study. In any case, to reliably measure changes in comprehension ability, a pre-post-test with episodes of exact difficulty should be used. It is possible that watching 24 episodes – spaced over eight months – might not have provided enough practise for students who had small prior experience with watching audio-visual input in English.

### **5.8.5. Summary**

Results from Study 1 confirm previous findings in the field regarding the higher efficiency of L1 subtitles over L2 captions for content comprehension at this level of proficiency, while corroborating the importance of vocabulary size when L2 captions are present – more demanding than subtitles for beginner-level learners. The study also suggests that explicit attention to target vocabulary items may have depleting effects on comprehension scores, which underlines the need to align the cognitive demands of tasks to learners' processing skills. Another valuable finding of the study concerns the interaction between item type and language of the on-screen text, suggesting that learners process textually explicit information and inferential information differently depending on the support they receive from the language available on the screen. The influence of item format and item type on comprehension also highlights the importance of taking into account item-related characteristics in the analysis. Finally, results corroborate the key role of lexical coverage as a strong predictor of comprehension, in line with findings from prior corpus-driven research and the few experimental studies existing in this area (e.g., Rodgers, 2013).





## Chapter 6. Vocabulary learning

---

### 6.1. Introduction

A rising number of studies have provided cumulative evidence in the past years that TV programmes and movies are efficient tools for incidental L2 vocabulary learning. Audio-visual input presents words in context with the additional support of images (Rodgers, 2013). It has also been found that words are more likely to reoccur in the same season of a TV programme (Rodgers & Webb, 2011), while viewing successive episodes of the same TV series also facilitates learning, as it allows viewers to accumulate background knowledge on the characters and the storyline – which in turn can help information processing.

Vocabulary learning can also be boosted by the addition of captions, as they aid aural form recognition, disambiguate unclear utterances, facilitate the identification of word boundaries, and help connecting the aural and written forms of words (e.g., Sydorenko, 2010; Winke et al., 2010). Captions, however, require a minimum L2 competence, as they cannot compensate for an excessive gap in proficiency (Guillory, 1998; Vanderplank, 2016b). Input beyond the linguistic competence of the learners may yield, therefore, poor language gains (Danan, 2004; Montero-Perez, Peters, Clarebout & Desmet, 2014). Studies have suggested that, for learners with limited L2 skills, subtitles might be more appropriate, as they can be processed automatically and provide access to the word meanings (Sydorenko, 2010). Findings from comparative studies, however, remain inconclusive, especially in the case of younger, less proficient learners.

Another way to optimize the effectiveness of vocabulary learning through TV programmes is the addition of explicit instruction. While the majority of studies in this

area have focused on incidental vocabulary learning, guessing meaning from context can be challenging due to the real-time nature of the input. Moreover, providing explicit access to the meaning – through pre-viewing activities or glossaries – may aid learners making an initial form-meaning connection. The scarce research in this area does suggest that having access to meaning while watching yields significantly higher vocabulary gains (Montero-Perez et al., 2018). While it has been proposed that pre-learning unknown topic-related words in a specific television programme may potentially improve comprehension and vocabulary learning (Webb, 2010b), to the author's knowledge, no studies have investigated the effect of pre-teaching lexical items prior to viewing to draw learners' attention to them, an enhancing technique that could be easily applied in the EFL classroom.

Study 2 aims at addressing these gaps by investigating vocabulary learning through the viewing of 24 episodes of a TV series over a period of 10 months. It will explore how learning of word forms and word meanings is affected by having access to captions or subtitles, and by having explicit instruction on target words (or not) prior to the viewing of the episodes, while also taking into account the effect that learners' L2 proficiency level might have on the learning outcomes. The relationship between vocabulary gains and general comprehension (from Study 1) will also be explored, in an attempt to yield a more complete picture of the data from both studies. As a follow-up, Study 2 also aims at assessing the long-term effects of the intervention on vocabulary retention rates, by testing participants' knowledge of the words learnt during the last term of the intervention eight months later.

## 6.2. Research questions

Study 2 is divided in two parts; the first one explores vocabulary gains through an extended exposure to a TV series, while the second investigates the long-term effects of the intervention. Specifically, the study will address the following research questions:

- Part One
1. To what extent can L2 vocabulary (form and meaning) be learnt through an extended exposure to TV series in the classroom?
  2. To what extent is vocabulary learning through TV series affected by:
    - a. Explicit focus on vocabulary (i.e. pre-teaching target lexical items or not)?
    - b. Language of the on-screen text (i.e. captions or subtitles)?
    - c. Learners' proficiency level?
  3. Does the percentage of vocabulary learning increase as more episodes are viewed?
  4. Is there a relationship between comprehension (Study 1) and vocabulary gains?
- Part Two
5. To what extent is the L2 vocabulary learnt through an extended exposure to TV series retained in the long term (i.e. eight months after the intervention)?

## **6.3. Part one: vocabulary learning**

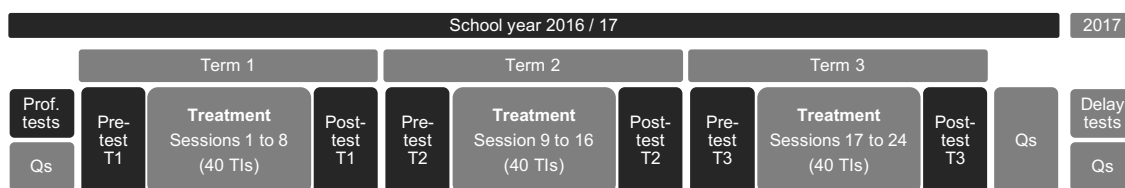
### **6.3.1. Participants**

As in Study 1, all Grade 8 students (N = 106; 65 females, 41 males) took part in the yearlong intervention, as the viewing sessions were embedded within the regular EFL subject. Only participants who had an attendance rate of 85% or more and completed all six pre-/post-tests were included in the analysis, leaving 80 participants (50 females, 30 males). According to the language of the on-screen text and whether they had instruction or had not, the groups were the following: captions focused (CF) (n = 22), captions non-focused (CNF) (n = 22), subtitles focused SF (n = 19), and subtitles non-focused (SNF) (n = 17). For part of the analysis, only participants who had completed the proficiency tests were included, leaving a final sample of 73 participants (45 female, 28 male).

### **6.3.2. Materials and procedure**

A complete description of the intervention materials can be found in Chapter 4. The instruments used for Study 2 are highlighted in a darker colour in Figure 6.1 below. Three sets of pre- and post-tests – administered at the beginning and at the end of each school term – were used to measure learning of word forms and word meanings (see section 4.4.3.1), and the OPT test was used as a general measure of initial proficiency (see section 4.4.1.1). Participants were tested on a total of 120 target items (see section 4.3.5, and Appendix A.2).

Figure 6.1. Materials used in Study 2 (part one)



The two measures of vocabulary – word-form and word-meaning gains – were obtained from three sets of pre- and post-tests, which were identical in format. In these vocabulary tests, participants were presented with the TIs aurally, and they had to write them down (i.e. word form) and then provide a translation or short definition (i.e. word meaning).

Both form and meaning were scored dichotomously (1 or 0). For form, a point was given when the word was correctly spelled. Since it was possible that a participant could correctly identify a word aurally but was just not able to transcribe it correctly, it was first considered the possibility to give partial points for words with spelling mistakes. However, the large variability of transcriptions per item made it nearly impossible to establish a clear cut for partially acceptable written words. For instance, the target item ‘bakery’ was transcribed with a wide variety of spellings, such as *backery*, *bakeri*, *vakery*, *baquery* or *beiqueri*. For an English speaker, the option *baykeri* might deserve partial credit, since the changes made would not affect how it is pronounced. The spelling *veiqueri*, on the other hand, which may look completely unidentifiable by a native speaker, sounds almost exactly as the word ‘bakery’ if you read it with Spanish pronunciation (*bakery* ['beikəri] vs. *veiqueri* ['beikeri]). This added a second problem, which was to decide whether proximity to the phonological form was to be determined based on the L1 or the L2. In view of this, and aware of its limitations, a strict spelling-based criterion was adopted, as it was considered it was a reliable and unbiased scoring method. For word meaning, translations and short definitions were scored by two raters. There was an interrater

reliability of 94.5%, and disagreement cases were discussed until an agreement was reached. A list of the accepted options was elaborated to keep track of the correct options, and ensure that the same correction criteria were applied from pre- to post-test.

A word was considered learnt when it was unknown in the pre-test and known in the post-test. Words known in both pre- and post-test were considered known but not learnt. Relative gains were calculated at item level following the formula used in previous studies (Horst, Cobb & Meara, 1988; Peters & Webb, 2018; Rodgers, 2013):

$$\text{Relative gains} = \frac{\text{number of learnt TIs}}{\text{total number of TIs} - \text{number of known TIs}} \times 100$$

For those participants who had missed a viewing session<sup>19</sup>, the TIs corresponding to that missed session(s) were not taken into account when calculating the percentage of relative vocabulary gains.

### 6.3.3. Preliminary analysis

Table 6.1 shows the descriptive statistics for the two proficiency measures: OPT and X\_Lex. At the time the intervention started, there were no significant differences between groups in terms of proficiency ( $F(3,69) = 1.280, p = .288$ ) nor vocabulary size ( $F(3,64) = .816; p = .490$ ).

---

<sup>19</sup> As stated above, participants with less than 85% attendance were excluded from the data, but there were a few participants who missed up to 3 viewing sessions (out of 24) who were included.

Table 6.1. General proficiency and Vocabulary size descriptive data

Group	General proficiency		Vocabulary size	
	n	OPT scores	n	X_Lex score
CF	21	99.71 (14.05)	17	1971 (547)
SF	18	92.06 (11.43)	16	2097 (332)
CNF	20	92.75 (15.04)	19	1992 (601)
SNF	14	93.43 (15.33)	16	1825 (434)
Average		94.71 (14.08)		1972 (494)

As in Study 1, participants were distributed into three proficiency levels according to the CEFR framework: Pre-A (n = 24), A1 (n = 35) and A2/B1 (n = 14). There were no significant differences in distribution amongst experimental groups ( $F(3,69) = .747, p = .528$ ). Because there were fewer participants with a valid vocabulary size score than with a proficiency score, in Study 2 OPT scores were used as the measure of proficiency<sup>20</sup>.

Results of the vocabulary pre- and post-tests are shown in Table 6.2 (with standard deviations in brackets). The table displays the total number of known word forms and word meanings (out of 40) at the start and by the end of Term 1 (T1), Term 2 (T2) and Term 3 (T3). Overall, prior to the intervention, participants knew an average of 22.85 word forms and 7.08 word meanings out of the total 120 TIs selected for the intervention. To assess whether the four experimental groups had similar level of knowledge of the target items at the beginning of each term, a series of ANOVAs were run. Results showed that there were no significant differences between the groups at

<sup>20</sup> Compared to Study 1, for Study 2 participants had to have completed the six vocabulary tests used to assess learning gains to be included in the analysis, besides the two proficiency tests. Participants who had all eight measures were only 61. To be able to maintain the highest number of participants, the OPT test (N=73) was chosen over the X\_Lex test (N=68).



any of the three pre-testing times, in either form (T1:  $F(3,76) = .347, p = .791$ ; T2:  $F(3,76) = 1.398, p = .250$ ; T3:  $F(3,76) = .505, p = .680$ ) nor meaning (T1:  $F(3,76) = .689, p = .562$ ; T2:  $F(3,76) = 1.020, p = .389$ ; T3:  $F(3,76) = .529, p = .664$ ).

Wilcoxon Signed Ranks tests also showed that differences between pre- and post-test were significant in all three terms for both form and meaning (for all six pairs,  $p < .001$ ) and for each experimental group (ranging from  $p < .001$  to  $p = .021$ ). The measure of relative gains used for the main analysis is the average relative gains across three terms, that is, the percentage of words learnt from pre- to post-tests out of the 120 TIs included in the intervention.

Table 6.2. Total scores in the pre- and post-tests for form and meaning

		Word form			Word meaning		
		T1	T2	T3	T1	T2	T3
<b>CF</b>	<b>Pre</b>	4.27 (4.84)	12.05 (6.82)	10.14 (7.14)	1.23 (2.33)	3.82 (3.40)	3.27 (4.63)
	<b>Post</b>	13.36 (8.11)	19.50 (9.40)	17.27 (9.62)	5.18 (4.58)	9.77 (6.95)	8.09 (7.16)
<b>SF</b>	<b>Pre</b>	3.16 (2.19)	8.79 (3.90)	7.79 (4.94)	0.74 (0.93)	2.68 (2.06)	1.68 (2.65)
	<b>Post</b>	9.68 (5.68)	15.42 (6.00)	12.16 (7.22)	3.11 (2.30)	6.11 (3.92)	4.53 (4.10)
<b>CNF</b>	<b>Pre</b>	4.05 (4.00)	9.18 (5.37)	9.14 (6.09)	0.59 (1.33)	3.32 (3.46)	2.55 (4.39)
	<b>Post</b>	6.23 (4.92)	12.73 (5.78)	12.36 (7.09)	1.73 (3.04)	5.59 (4.54)	5.23 (6.52)
<b>SNF</b>	<b>Pre</b>	3.94 (2.56)	10.12 (6.02)	8.71 (5.83)	0.94 (0.97)	4.59 (3.66)	2.88 (4.51)
	<b>Post</b>	8.00 (4.29)	13.65 (8.12)	11.00 (6.78)	1.65 (1.92)	7.41 (6.20)	5.71 (5.99)
<b>Mean</b>	<b>Pre</b>	3.88 (3.61)	10.08 (5.70)	9.00 (6.14)	0.88 (1.54)	3.58 (3.40)	2.61 (4.11)
	<b>Post</b>	9.39 (6.53)	15.45 (7.82)	13.38 (8.08)	2.99 (3.66)	7.25 (5.70)	5.95 (6.16)

### 6.3.4. Results

#### 6.3.4.1. Learning vocabulary through viewing TV series

The first research question aimed at investigating the extent to which participants could learn L2 vocabulary through an extended exposure to TV series in the classroom. For this analysis, the mean percentage of gains for form and meaning across the intervention was calculated to obtain a global score. Table 6.3 shows the mean percentage of relative gains in form and meaning for the 120 targeted lexical items.

Table 6.3. Percentage of relative gains for form and meaning out of 120 target items

		Percentage of relative gains					
Group	n	Form			Meaning		
		Min	Max	Mean (SD)	Min	Max	Mean (SD)
CF	22	5.40	62.18	<b>30.25 (16.45)</b>	0.83	32.11	<b>14.63 (10.19)</b>
SF	19	6.72	41.56	<b>21.53 (11.16)</b>	0.00	20.00	<b>8.45 (6.36)</b>
CNF	22	3.77	29.93	<b>13.90 (6.34)</b>	0.83	25.75	<b>6.72 (6.41)</b>
SNF	17	1.90	31.68	<b>14.32 (8.77)</b>	0.00	24.36	<b>8.56 (7.49)</b>
All	80	1.90	62.18	<b>20.30 (13.18)</b>	0.00	32.11	<b>9.70 (8.31)</b>

For form learning, overall, the two focused groups (who had been taught the TIs in advance) performed better than the non-focused groups, with the captions group (CF) – with simultaneous exposure to both L2 sound and text – outperforming the subtitles group (SF). The least successful groups were the two non-focused groups (SNF and CNF), who performed similarly independently of the language of the on-screen text. For meaning learning, the CF group outperformed again the other three conditions. The two subtitles groups (SF and SNF) – with access to the L1 translations – performed similarly, no matter the instruction condition, and the CNF group scored slightly below them.

It is interesting to note, however, that the range of gains within the experimental groups is considerable. For example, the most successful group (i.e. CF) had 30.10% gains in form, and 14.64% gains in meaning, which indicates that, on average, participants in that group learnt approximately 36 word forms and 18 word meanings. However, in that group the most successful participants learning rate was almost double the average, with gains of 62.18% in form and 32.11% in meaning, while the least successful learners had gains of 5.40% in form and 0.83% in meaning.

Differences among the groups in word form gains were first explored by means of a Welch's ANOVA (a Levene's test showed that variances were unequal:  $F(3,76) = 2.774$ ,  $p = .047$ ). The test showed that there was a statistically significant difference between groups ( $F(3,76) = 7.714$ ,  $p < .001$ ,  $\omega^2 = .199$ ) and that approximately 20% of the total variance in the percentage of form-learning gains was accounted for by the experimental group. A Tamhane's T2 post hoc test revealed that the difference was significant between the CF and CNF groups ( $p = .001$ ) as well as between the CF and SNF groups ( $p = .003$ ). For meaning recall, a one-way ANOVA showed that there were also significant differences between the experimental groups ( $F(3,76) = 3.301$ ,  $p = .024$ ), and a Tukey HSD post-hoc test revealed that differences were only significant between CF and CNF groups ( $p = .023$ ).

#### *6.3.4.2. Mediating factors: on-screen text, type of instruction and proficiency*

The second research question aimed at investigating to what extent language of the on-screen text, type of instruction and learners' proficiency explained the differences in vocabulary learning observed between the four groups. Two generalized linear models (GLMs) were run to evaluate the influence of these three factors on the two vocabulary outcome measures: form learning and meaning learning.

A GLM was first calculated with the percentage of word-forms learnt as the dependent variable, and on-screen text language (captions or subtitles), type of instruction (focused or non-focused), and proficiency (Pre-A, A1 or A2/B1) as fixed effects. Non-significant interactions were removed from the model, leaving only main effects. Table 6.4 below presents the final GLM model, and Table 6.5 reports the pairwise contrast between categorical factors.

Table 6.4. GLM results for word-form learning

Terms	Coefficient	SD	t	Sig	95% CI <sup>a</sup>	
					Lower	Upper
Intercept	26.155	3.2488	8.051	<.001	19.672	32.638
Language (S)	-3.374	2.4975	-1.351	.181	-8.358	1.609
Language (C)	0 <sup>b</sup>	.	.	.	.	.
Instruction (F)	11.785	2.4748	4.762	<.001	6.847	16.724
Instruction (NF)	0 <sup>b</sup>	.	.	.	.	.
Proficiency (Pre-A)	-16.497	3.5600	-4.634	<.001	-23.601	-9.394
Proficiency (A1)	-11.469	3.3313	-3.443	.001	-18.117	-4.822
Proficiency (A2/B1)	0 <sup>b</sup>	.	.	.	.	.

<sup>a</sup> Confidence interval

<sup>b</sup> Coefficient is set to zero because is redundant

Table 6.5. GLM results for word-form learning; pairwise contrasts

	Mean (SE)	M Diff (SE)	df	F	Sig.
Captions	22.72 (1.70)				
Subtitles	19.35 (1.95)	3.37 (2.50)	1, 68	1.825	.181
Focused	26.93 (1.75)				
Non-Focused	15.14 (1.89)	11.78 (2.47)	1, 68	22.679	<.001
Pre-A <sup>a</sup>	13.86 (2.15)	<sup>a-b</sup> 5.03 (2.79)			
A1 <sup>b</sup>	18.89 (1.79)	<sup>b-c</sup> 11.46 (3.33)	2, 68	10.894	<.001
A2/B1 <sup>c</sup>	30.36 (2.84)	<sup>a-c</sup> 16.50 (3.56)			

The final model revealed that there was a main effect of type of instruction and proficiency, but that language condition did not predict word form learning. Results showed that participants in the focused condition scored significantly higher ( $p < .001$ ) than their classmates by about 11.78%. Gains in form also depended significantly on participants' proficiency level ( $p = .001$ ), with the most proficient students (A2/B1 level) scoring 11.46% higher than A1-level students and 16.50% higher than Pre-A-level students. Pairwise contrasts revealed that differences were significant between Pre-A and A2/B1 ( $p < .001$ ) and A1 and A2/B1 ( $p = .002$ ), but only marginal between Pre-A and A1 level ( $p = .076$ ).

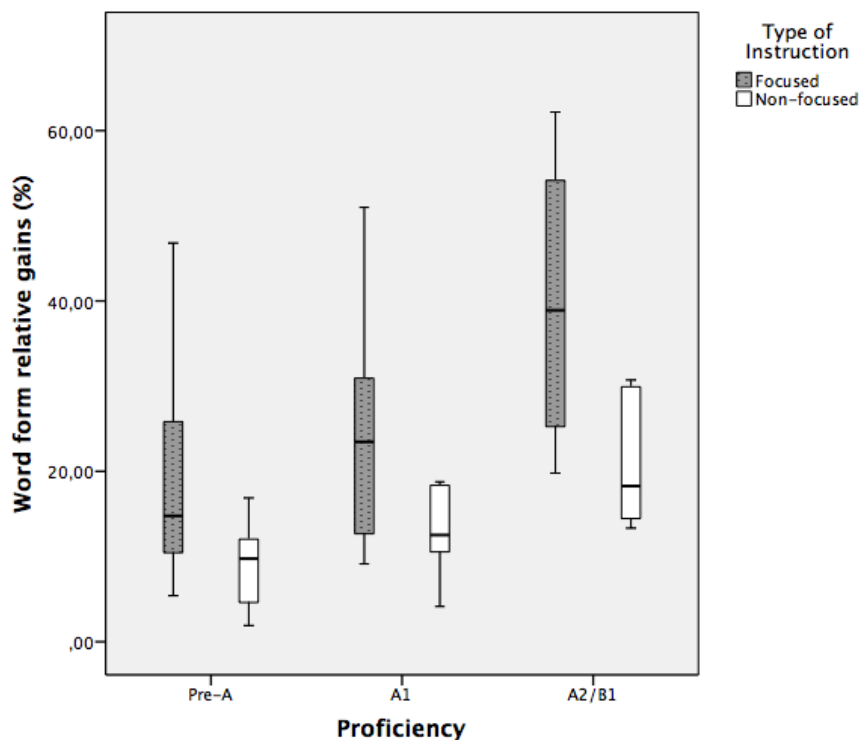
While no interaction emerged between instruction and proficiency, the effect of learners' proficiency in each instruction group was further explored. For the analysis, the CF and SF groups were jointly considered *focused*, and CNF and SNF were *non-focused* (for no significant effect of on-screen language was found). Table 6.6 shows the percentage of relative gains per type of instruction and proficiency group, with standard error (SE) in brackets.

Table 6.6. Relative gains in form and meaning per type of instruction and proficiency group

Type of Instruction	Proficiency	n	Percentage of relative gains Mean (SE)	
			Form	Meaning
Focused	Pre-A	12	18.68 (3.05)	7.70 (2.09)
	A1	19	24.23 (2.43)	10.65 (1.66)
	A2/B1	8	39.82 (3.74)	20.08 (2.56)
Non-focused	Pre-A	12	9.05 (3.05)	2.83 (2.09)
	A1	16	14.19 (2.64)	7.14 (1.81)
	A2/B1	6	20.83 (4.32)	14.65 (2.95)

It was found that there were significant differences between types of instruction when comparing each proficiency level against its counterpart, with the focused group significantly outperforming the non-focused group at the Pre-A ( $F(1,67) = 4.975, p = .029$ ), A1 ( $F(1,67) = 7.833, p = .007$ ) and A2/B1 ( $F(1,67) = 11.066, p = .001$ ) levels. Within the focused group itself, results indicated that differences were also significant between the three levels of proficiency ( $F(2,67) = 9.970, p < .001$ ), and pairwise comparisons showed that significant differences were found between the most proficient group (A2/B1 level) and both Pre-A ( $p < .001$ ) and A1 levels ( $p = .002$ ). In contrast, in the non-focused group differences were marginally significant ( $F(2,67) = 2.538, p = .087$ ), and did not reach significance between any of the groups. Figure 6.2 shows estimated marginal means per focused and non-focused groups when participants were divided by proficiency levels (Pre-A, A1 and A2/B1).

Figure 6.2. Estimated marginal means for word-form learning: proficiency and instruction



A second GLM was calculated with percentage of word-meanings learnt as the dependent variable, and language (captions or subtitles), type of instruction (focused or non-focused) and proficiency (Pre-A, A1 or A2/B1) as fixed effects. Again, non-significant interactions were removed from the model, leaving only the main effects. Table 6.7 and Table 6.8 below present the results of the final model.

Table 6.7. GLM results for word-meaning learning

Terms	Coefficient	SD	t	Sig	95% CI <sup>a</sup>	
					Lower	Upper
Intercept	15.754	2.2100	7.129	.000	11.344	20.164
Language (S)	-1.433	1.6989	-.844	.402	-4.824	1.957
Language (C)	0 <sup>b</sup>	.	.	.	.	.
Instruction (F)	4.398	1.6835	2.613	.011	1.039	7.758
Instruction (NF)	0 <sup>b</sup>	.	.	.	.	.
Proficiency (Pre-A)	-11.969	2.4217	-4.942	<.001	-16.801	-7.136
Proficiency (A1)	-8.483	2.2661	-3.743	<.001	-13.005	-3.961
Proficiency (A2/B1)	0 <sup>b</sup>	.	.	.	.	.

<sup>a</sup> Confidence interval

<sup>b</sup> Coefficient is set to zero because is redundant

Table 6.8. GLM results for word-meaning learning: pairwise contrasts

	Mean (SE)	M Diff (SE)	df	F	Sig.
Captions	11.14 (1.15)				
Subtitles	9.70 (1.33)	1.43 (1.70)	1, 68	.712	.402
Focused	12.62 (1.19)				
Non-Focused	8.22 (1.29)	4.40 (1.68)	1, 68	6.826	.011
Pre-A <sup>a</sup>	5.27 (1.46)	<sup>a-b</sup> 3.49 (1.90)			
A1 <sup>b</sup>	8.75 (1.22)	<sup>b-c</sup> 8.48 (2.26)	2, 68	12.373	<.001
A2/B1 <sup>c</sup>	17.24 (1.93)	<sup>a-c</sup> 11.97 (2.42)			

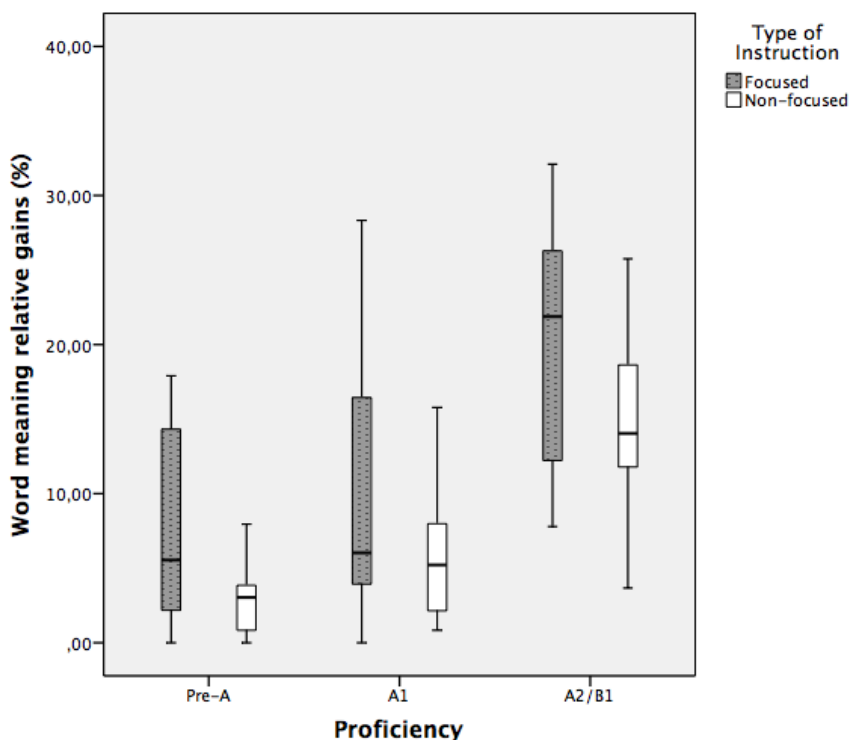
Similar to what had been found for form gains, the final model showed a main effect of type of instruction and proficiency, but no main effect of language of the on-screen text. Participants in the focused condition scored significantly higher ( $p = .020$ ) than participants in the non-focused groups by 4.4%. Gains in meaning also depended significantly on participants' proficiency level ( $p < .000$ ), with the most proficient students (A2/B1 level) scoring 11.97% higher than the least proficient group (Pre-A).

Once more, the relationship between instruction and proficiency was examined (see Table 6.6 above). In contrast with form, for meaning learning there were no significant differences between instruction groups at the Pre-A ( $F(1,67) = 2.715, p = .104$ ), the A1 ( $F(1,67) = 2.038, p = .158$ ) nor the A2/B1 level ( $F(1,67) = 1.933, p = .169$ ), although learners in the focused group had consistently higher gains in meaning than learners in the non-focused group.

Differences between proficiency levels were significant, however, within each instruction group (focused:  $F(2,67) = 7.410, p = .001$ ; non-focused:  $F(2,67) = 5.351, p = .007$ ). Pairwise comparisons revealed that in both conditions it was the more advanced group (A2/B1 level) the one that significantly outperformed the other two in meaning recall. In the focused group differences were found between the A2/B1 level and both the Pre-A ( $p = .001$ ) and the A1 levels ( $p = .006$ ). In the non-focused group, they only found between the A2/B1 and the Pre-A level ( $p = .005$ ), while they were only marginal between A2/B1 and A1 level ( $p = .067$ ). Figure 6.3 shows estimated marginal means per type of instruction when participants were divided by proficiency level.



Figure 6.3. Estimated marginal means for word-meaning learning: proficiency and instruction



6.3.4.3. Learning vocabulary through multiple, successive episodes

The third research question explored – similar to Study 1 – whether an extensive exposure could increase gains over time, as learners were watching successive episodes and were getting used to the TV series. Table 6.9 displays the percentage of relative gains in form and meaning learning (SD in brackets) per term and experimental condition.

Table 6.9. Percentage of gains in word-form and word-meaning per term and experimental condition

	n	Term 1		Term 2		Term 3	
		Form	Meaning	Form	Meaning	Form	Meaning
CF	22	27.59 (16.64)	10.69 (8.32)	33.19 (18.09)	18.58 (13.23)	29.97 (18.12)	14.63 (11.30)
SF	19	19.86 (11.62)	6.82 (6.94)	26.08 (12.81)	10.07 (7.07)	18.66 (12.73)	8.48 (7.88)
CNF	22	9.28 (7.30)	3.31 (6.83)	16.99 (7.32)	8.03 (6.04)	15.44 (9.79)	8.81 (7.88)
SNF	17	13.17 (8.26)	2.14 (3.38)	18.24 (13.87)	11.44 (11.23)	11.55 (8.28)	12.09 (12.79)
Mean		17.66 (13.60)	65.92 (7.46)	23.87 (12.93)	12.14 (10.51)	19.37 (14.58)	11.03 (10.55)

As can be observed, there was an increase in the percentage of word-form learning from T1 to T2 across all four conditions, but not in T3 – when gains decreased. For word-meaning learning, gains were also higher in T2, and remained similar in T3. GLMs were run again, but this time per each term separately. Results were similar for the three terms, and similar to the overall results: type of instruction and proficiency level emerged as significant predictors of both form and meaning gains, while language of the on-screen text did not contribute to the models. No interactions emerged in any of the six models.

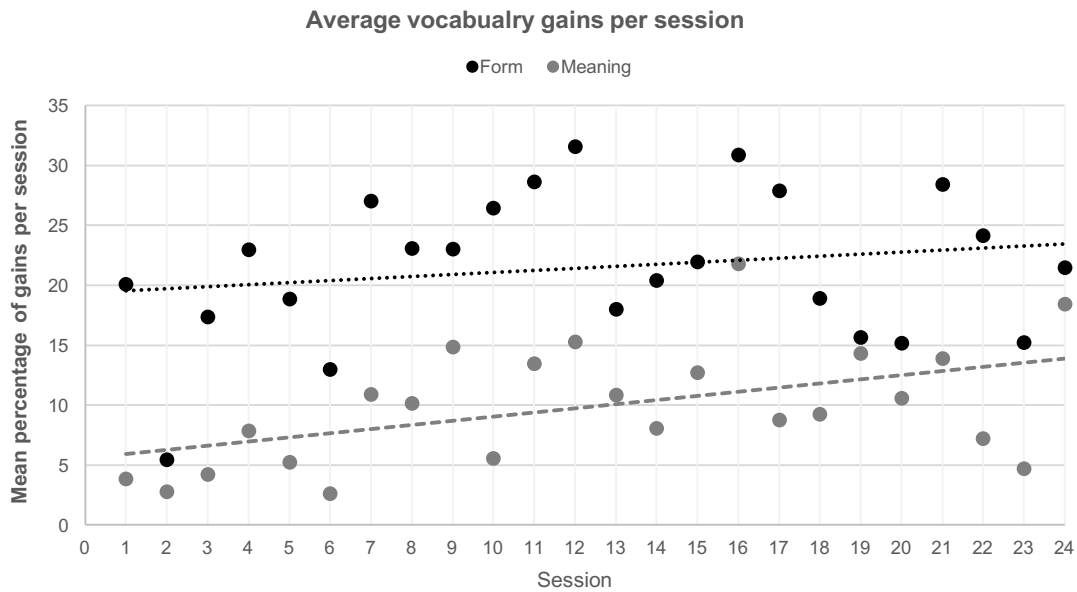
There were only two exceptions in T3, in which language condition also emerged as a predictor for form learning (with the captions group outperforming the subtitles groups as usual, but this time significantly), and instruction condition did not significantly predict gains in meaning (although the focused groups still had higher gains than their counterparts). Table 6.10 provides a summary of the six models.

Table 6.10. Summary of results from GLMs for form and meaning, by term

Factor	Condition	Form		Meaning		
		Contrast estimate	Sig.	Contrast estimate	Sig.	
T1	Language	C > S	1.14	.586	1.70	.256
	Instruction	F > NF	12.50	< .001	5.61	< .001
	Proficiency	PrA> A1 > A2/B1	4.10 / 12.11	< .001	2.38 / 7.15	< .001
T2	Language	C > S	1.98	.507	1.24	.574
	Instruction	F > NF	12.41	< .001	5.00	.025
	Proficiency	PrA> A1 > A2/B1	4.48 / 12.76	< .001	4.10 / 10.44	< .001
T3	Language	C > S	6.70	.023	1.36	.555
	Instruction	F > NF	10.44	< .001	2.58	.259
	Proficiency	PrA> A1 > A2/B1	6.50 / 9.53	.001	3.99 / 7.86	.002

Learning from the first to the last episode was explored last by looking at the percentage of gains per item (i.e. the percentage of participants who learnt a particular item) and the mean percentage of gains per episode (i.e. the average gains of the five TIs in the episode). Figure 6.4 shows the changes over time in word-form (in black) and word-meaning learning (in grey), as well as the corresponding fitted lines. Results from a linear regression with session (i.e. from 1 to 24) as a predictor showed that number of session did not predict word-form learning ( $F(1, 118) = .749$ ;  $p = .389$ ,  $R^2 = .006$ ), but it accounted for 5% of the variance of gains in meaning ( $F(1, 118) = 6.197$ ;  $p = .014$ ,  $R^2 = .050$ ). The variability of gains per item within the episode suggests that difference in learning may be explained by other more relevant word-related factors, such as frequency of encounters or distribution of the encounters in the input. This will be further explored in Study 3 (see Chapter 7).

Figure 6.4. Average percentage of gains in form and meaning per session



#### 6.3.4.4. Comprehension scores and vocabulary gains

The study's fourth research question aimed at exploring the relationship between vocabulary gains and the overall comprehension of the episodes obtained in Study 1 (see Chapter 5).

A Spearman's rank-order correlation was first computed to assess the relationship between comprehension and the relative gains in word form. There was a small, positive correlation between comprehension and word-form learning, which was statistically significant ( $r_s(80) = .289, p = .009$ )<sup>21</sup>. A second correlation was run for gains in meaning. In this case, a medium, positive correlation was found between comprehension and word-meaning learning, which was statistically significant ( $r_s(80) = .381, p < .001$ ). This relationship was further explored by running a series of correlations within each

<sup>21</sup> For correlation coefficients, the rule of thumb followed in the present study is that values close to .25 are considered small, .40 medium, and .60 large, as proposed by Plomsky and Oswald (2014).

experimental group. Results showed that the relationship between vocabulary gains and general comprehension seemed to be stronger in the groups with access to captions – particularly in the CF group –, and for word-meaning learning, while it was only marginal in the case of the SNF group, as can be observed in Table 6.11. The relationship between comprehension and the two vocabulary variables (gains in form and meaning) is illustrated in Figure 6.5 and Figure 6.6.

Table 6.11. Correlations between word learning and comprehension

		<b>Comprehension</b>
CF (n = 22)	Form	.791** (<.001)
	Meaning	.772** (<.001)
CNF (n = 22)	Form	.632** (.002)
	Meaning	.688** (<.001)
SF (n = 22)	Form	.477* (.039)
	Meaning	.674** (.002)
SNF (n = 22)	Form	.435 (.081)
	Meaning	.458 (.065)

Figure 6.5. Word-form learning and comprehension

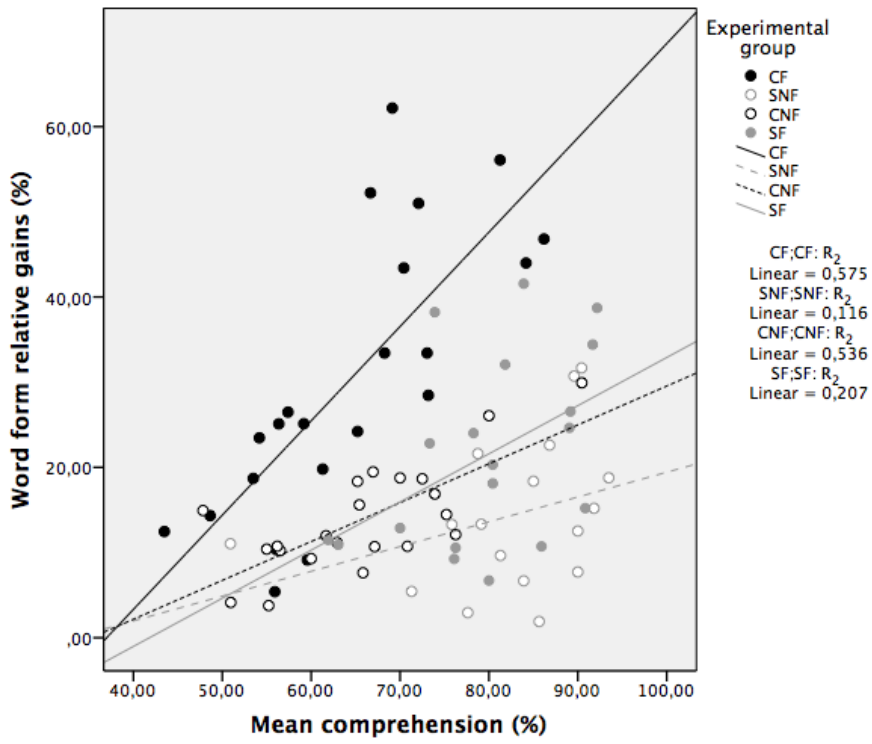
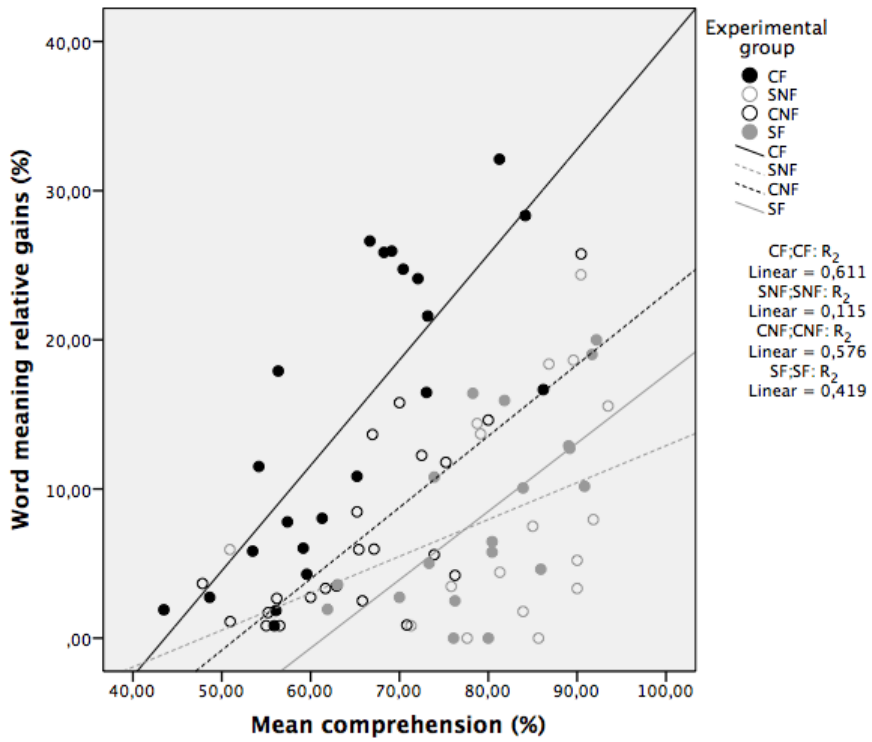


Figure 6.6. Word-meaning learning and comprehension



A regression analysis revealed that, for word-form learning, the percentage of comprehension significantly accounted for around 57.5% and 53.6% of the variance in the CF and CNF groups respectively, and marginally for the SF group (20.7%), while comprehension was not a significant predictor for the SNF condition. For word-meaning learning, the percentage of variation explained by comprehension was 61.1% for CF, 57.6% for CNF and 41.9% for the SF group, while it did not predict – again – gains in meaning for the SNF group (see Table 6.12).

Table 6.12. Results from regression analysis

	Group	R	R <sup>2</sup>	Adj R <sup>2</sup>	SE	R <sup>2</sup> Change	F Change	df1	df2	Sig. F Change
Form	CF	.758	<b>.575</b>	.554	10.989	.575	27.046	1	20	.000
	SF	.455	<b>.207</b>	.160	10.233	.207	4.426	1	17	.051
	CNF	.732	<b>.536</b>	.513	4.423	.536	23.136	1	20	.000
	SNF	.341	<b>.116</b>	.057	8.520	.116	1.968	1	15	.181
Meaning	CF	.781	<b>.611</b>	.591	6.513	.611	31.369	1	20	.000
	SF	.647	<b>.419</b>	.384	4.989	.419	12.244	1	17	.003
	CNF	.759	<b>.576</b>	.555	4.280	.576	27.188	1	20	.000
	SNF	.339	<b>.115</b>	.056	7.283	.115	1.946	1	15	.183

## 6.4. Part two: long term retention

In the first part of Study 2, it was established that learners could acquire new vocabulary through exposure to TV series, and that success depended on whether learners were pre-taught the target words in advance and also on their L2 proficiency level. The second part of Study 2 was designed as a follow-up of the intervention, with the aim of assessing whether the words that had been learnt were retained in the long term. This second part focuses on the 40 words targeted in the last term (i.e. T3), and seeks to explore if and how differently the TIs learnt are retained by learners in the different experimental conditions.

### **6.4.1. Participants**

Participants were the same Grade 8 students who had taken part in Study 1 and 2, and who were now – eight months after the intervention – in Grade 9. Since tests were administered within a regular EFL class, all Grade 9 students took the tests of the study, but only those who were valid in T3 (i.e. had an 85% attendance or more and had completed the T3's pre- and post-tests) were included in the analysis. In total, there were 82 participants (52 female, 30 male). For part of the analysis, only participants who had also completed the proficiency tests at the beginning of Grade 8 were included.

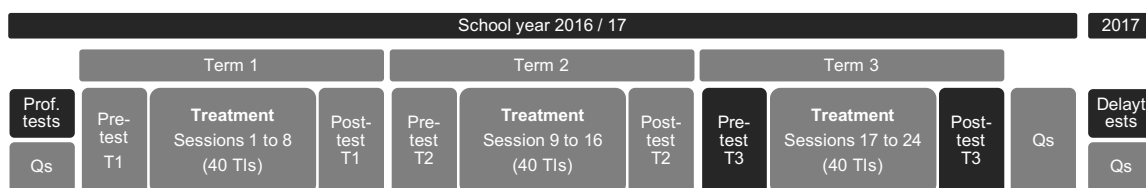
### **6.4.2. Materials and procedure**

A complete description of the materials specific to this part of the study can be found in Chapter 4. The instruments for the analysis of long-term vocabulary retention are highlighted in a darker colour in Figure 6.7 below. Namely, the tests used were: (a) the OPT test as a measure of proficiency (section 4.4.1.1); (b) the T3's pre- and post-test as baseline for words learnt during that term (section 4.4.1.3); and (c) two delayed post-tests as a measure of vocabulary retention, which assessed the participants' knowledge of the same 40TIs appearing in T3 (section 4.4.1.4).

The first delayed post-test had the same format as the pre- and post-tests: participants had to listen to the TIs, write down the word in English, and then provide a translation in Catalan or Spanish. The second delayed post-test included the same 40 TIs, but it already provided the written form of the words (instead of the aural prompt), and participants only had to write down the translation. This was administered once their first test sheets had already been collected.



Figure 6.7. Materials used in Study 2 (part two)



The two delayed post-tests items were scored dichotomously (0 or 1), following the same procedure as in the other pre- and post-tests. For word form, the spelling-based scoring criterion was applied. For word meaning, translations were scored by two raters, who took into account the list of the accepted and rejected options that resulted from the correction of the pre- and post-test in T3.

Because the main interest of the study was assessing the percentage of TIs *learnt* during the intervention that were *retained* in the long term, results in the delayed post-test were coded taking into account the pre-, post- and delayed-test scores. According to those, there were eight different outcomes depending on whether a TI was unknown, learnt or forgotten from pre- to post-test, and from post-test to delayed post-test, as shown in Table 6.13. For the following analysis, only categories 3 and 4 were used – words *learnt* in the intervention that were either *retained* or *forgotten* in the delayed test.

Table 6.13. Scoring in delayed post-test

Pre-test T3	Post-test T3	Delayed test	Outcome	Coding
0	0	0	Unknown	1
0	0	1	Unknown & learnt	2
0	1	1	Learnt & retained	3
0	1	0	Learnt & forgotten	4
1	0	0	Forgotten	5
1	0	1	Forgotten & learnt	6
1	1	1	Known & retained	7
1	1	0	Known & forgotten	8

### 6.4.3. Preliminary analysis

Table 6.14 reports again (for clarity reasons) the percentage of vocabulary gains in T3. As discussed above, in T3 differences between groups were significant for form but not for meaning, although the same tendencies found in the other two terms could be observed, with the focused groups outperforming the non-focused ones and learners in the captions condition performing better than learners in the subtitles condition.

Table 6.14. Percentage of relative gains in form and meaning in T3

	n	Term 3	
		Form	Meaning
<b>CF</b>	22	29.97 (18.12)	14.63 (11.30)
<b>SF</b>	19	18.66 (12.73)	8.48 (7.88)
<b>CNF</b>	22	15.44 (9.79)	8.81 (7.88)
<b>SNF</b>	17	11.55 (8.28)	12.09 (12.79)
<b>Mean</b>		19.37 (14.58)	11.03 (10.55)

Figure 6.6 and Figure 6.7 show the absolute test scores (out of 40) for both form and meaning by experimental group and each testing time (pre-, post- and delayed post-test). As can be observed, participants scored similarly in the post- and delayed post-tests for both form and meaning, with overall slightly higher scores for the non-focused groups. For meaning recall, scores for the written input test were consistently higher than the aural test.

Figure 6.8. Results for word forms in the pre-, post- and delayed-post-test

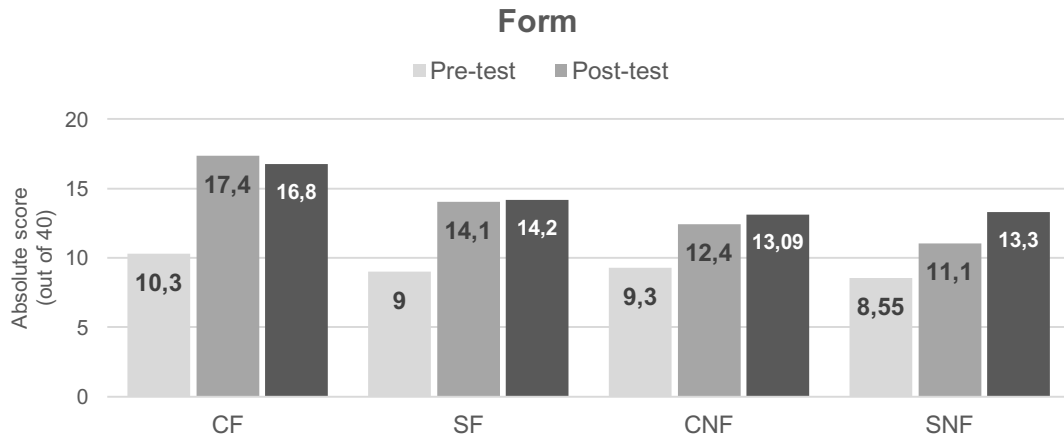
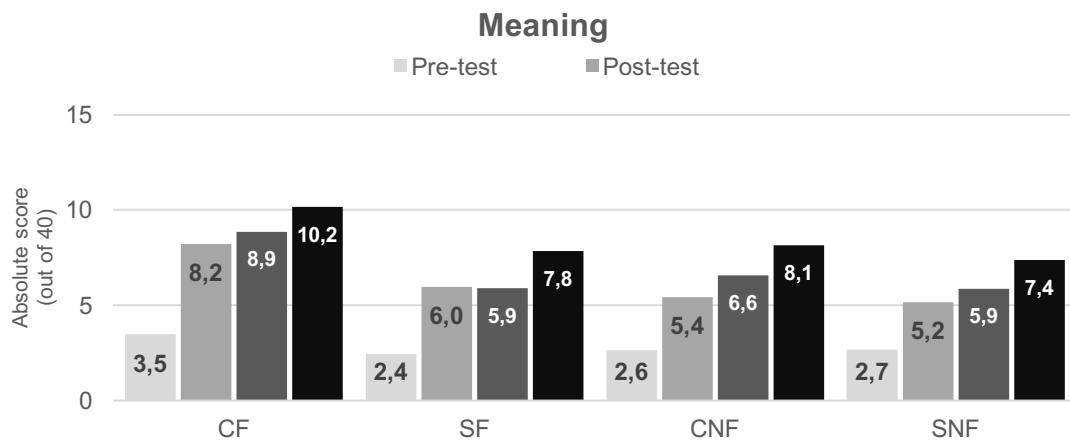


Figure 6.9 .Results for word meaning in the pre-, post- and delayed-post-test



## 6.4.4. Results

### 6.4.4.1. Vocabulary long-term retention

The second part of Study 2 aimed at unveiling the percentage of TIs learnt during the intervention that were retained eight months later. Differences between pre-, post- and delayed post-test scores were first examined.

Results from a series of Wilcoxon Signed-Ranks test indicated that, for both word-form and word-meaning learning, there was a significant difference from pre- to post-test for all four experimental groups. From word-form learning, there were no differences between post- and delayed post-test (except in the SNF group, who did better in the delayed post-test). Similarly, for word-meaning learning there were no difference between the post- and delayed post-test with aural input (except in the CNF group), but results were significantly higher for all conditions in the delayed post-test with written input.

		CF	SF	CNF	SNF
<b>Form</b>	Pre - Post	-3.832 <sup>a</sup>	-3.643 <sup>a</sup>	-3.724 <sup>a</sup>	-3.207 <sup>a</sup>
		<.001	<.001	.001	<.001
	Post - Delay	-1.032 <sup>b</sup>	-.156 <sup>a</sup>	-.658 <sup>a</sup>	-2.755 <sup>a</sup>
		.302	.876	.511	.006
<b>Meaning</b>	Pre - Post	-3.631 <sup>a</sup>	-3.212 <sup>a</sup>	-3.527 <sup>a</sup>	-3.426 <sup>a</sup>
		<.001	.001	<.001	.001
	Post - Delay A <sup>c</sup>	-1.332 <sup>a</sup>	-.154 <sup>b</sup>	-1.990 <sup>a</sup>	-1.375 <sup>a</sup>
		.183	.878	.047	.169
	Post - Delay W <sup>d</sup>	-3.366 <sup>a</sup>	-2.862 <sup>a</sup>	-3.688 <sup>a</sup>	-3.114 <sup>a</sup>
	.001	.004	<.001	.002	
Delay A - W	-3.090 <sup>a</sup>	-3.155 <sup>a</sup>	-3.493 <sup>a</sup>	-3.140 <sup>a</sup>	
	.002	.002	<.001	.002	

<sup>a</sup> Based on negative ranks

<sup>b</sup> Based on positive ranks

<sup>c</sup> Delayed post-test with aural input

<sup>d</sup> Delayed post-test with written input

Table 6.15 shows the percentage of words forms and word meanings learnt, and what percentage of those were retained and forgotten eight months later. For the meaning written-format test, note that the percentage of words learnt was assessed by aural-format tests, and comparison should be done with caution.

Table 6.15. Percentage of retained word forms and meanings

	n	FORM			MEANING Aural			MEANING Written		
		Learnt	Retained	Forgotten	Learnt	Retained	Forgotten	Learnt	Retained	Forgotten
<b>CF</b>	20	30.4%	63.6%	36.4%	14.5%	74.8%	25.2%	14.5%	73.5%	26.5%
<b>SF</b>	19	22.1%	61.9%	38.1%	10.7%	56.5%	43.5%	10.7%	81.2%	18.8%
<b>CNF</b>	23	15.1%	69.6%	30.4%	9.3%	62%	38%	9.3%	76.7%	23.3%
<b>SNF</b>	20	11.6%	63.8%	36.2%	11.9%	52.5%	47.5%	11.9%	66.8%	33.2%

Overall, learners retained between 62 and 72% of the word forms learnt, and between 52 and 75% of the word meanings in the aural-format test. A series of Welch ANOVAs were run to explore the variance in retention rates between the four conditions. Results showed that, for form retention, differences between experimental groups did not reach significance ( $F(3,75) = 40.681, p = .852$ ). Similarly, there were no significant differences in meaning retention in the aural input test ( $F(3,63) = 33.925, p = .148$ ) nor in the written input test ( $F(3,63) = 34.918, p = .603$ ). For form retention, the CNF group was the most successful, while for meaning retention it was the CF the group that had a higher retention rate.

When retention of meaning was assessed in the written format test, retention rates were higher, ranging from 67 to 81%, with the SF group displaying the highest percentage. A Wilcoxon signed-rank test showed that there was, overall, a significant

difference between the percentage of words retained in the aural test ( $M = 61.5$ ;  $SD = 34.5$ ) and the written test ( $M = 74.4$ ;  $SD = 29.1$ ) ( $Z = -3.744$ ;  $p < .000$ ), with the latter one reporting 12.9% higher meaning retention. When the analyses were run again by experimental group, however, this was only true for the two subtitles groups (SF:  $Z = -2.527$ ,  $p = .012$ ; SNF:  $Z = -2.371$ ,  $p = .018$ ). For the CNF group it was marginally significant ( $Z = -1.859$ ,  $p = .060$ ), while for the CF group the difference in percentage of retention did not reach statistical significance ( $Z = -1.187$ ,  $p = .235$ ).

Besides the words that had been retained (or forgotten) from post- to delayed post-test, there was a percentage of words that were unknown in the pre- and post-test but that had been learnt by the time the delayed post-test was administered. Regarding word-form, there were an average of 5.06 words learnt in the focused groups, and 7.05 in the non-focused groups. As for word-meaning, there were 5.84 and 5.41 words learnt in the focused and non-focused groups respectively, which rose to 7.92 and 8.47 words in the written-format meaning test. Figures 6.10, 6.11 and 6.12 show the distribution of the eight outcomes in the delayed post-test by type of instruction.

Figure 6.10. Delayed post-test results: word form

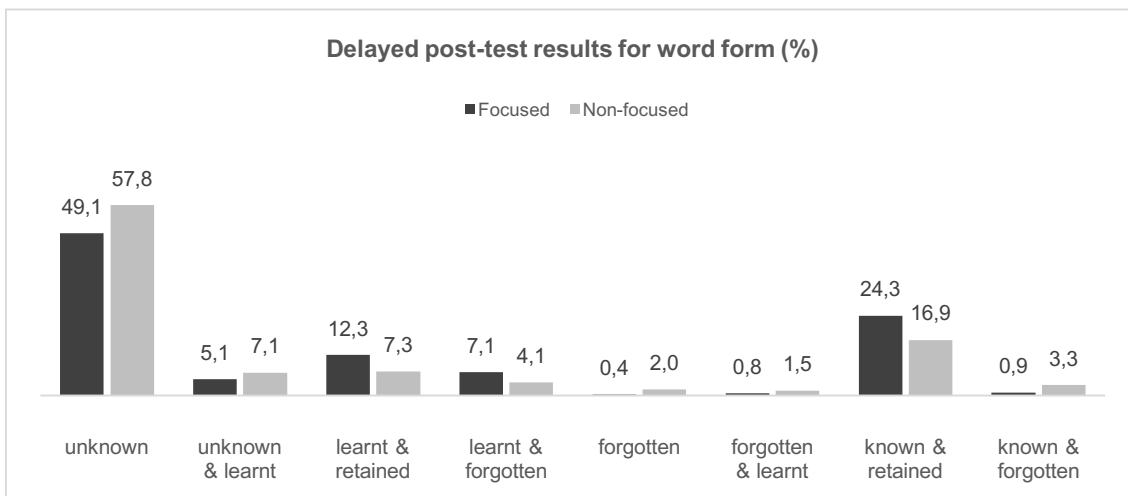


Figure 6.11. Delayed post-test results: word meaning (aural test)

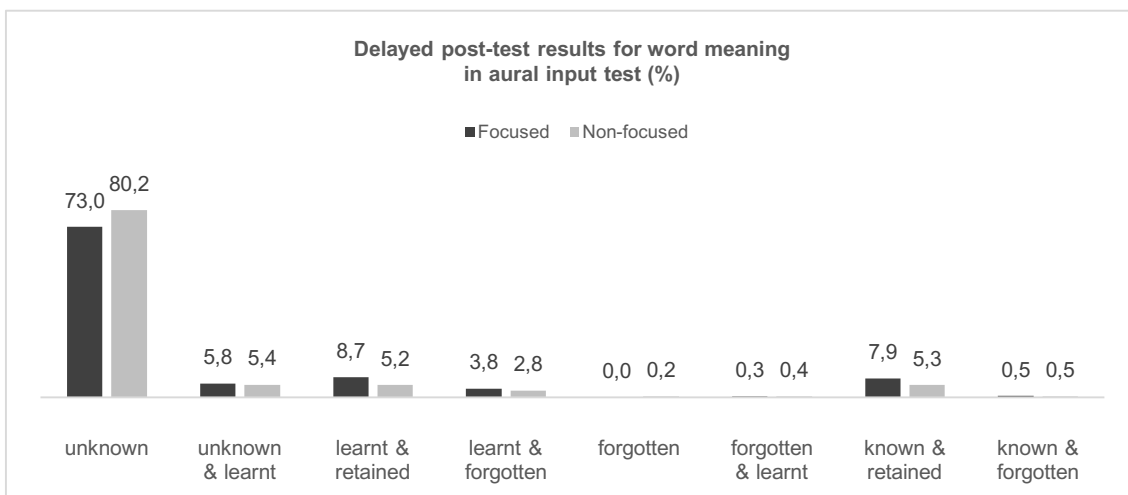
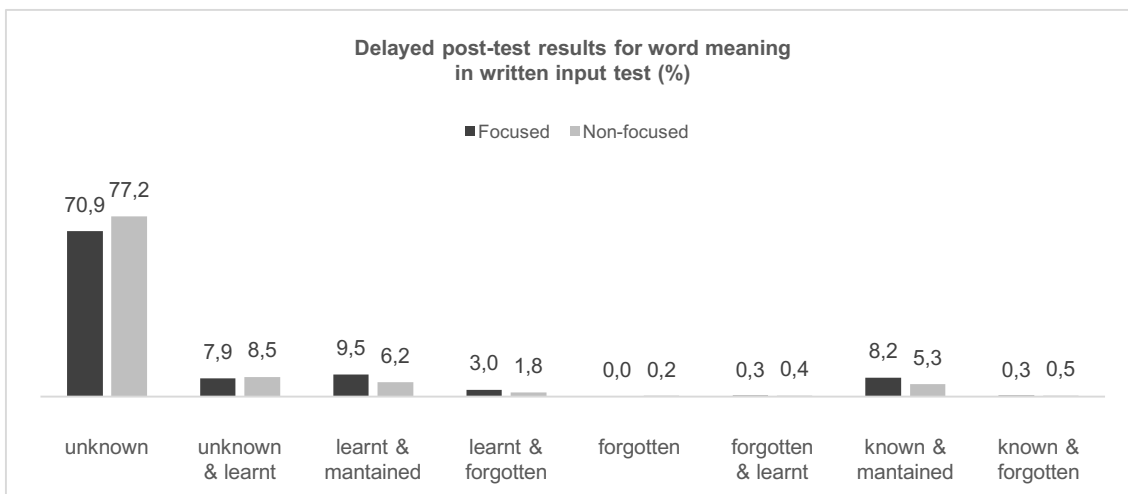


Figure 6.12. Delayed post-test results: word meaning (written test)



## 6.5. Summary of findings

The main findings from Study 2 can be summarized as follows:

1. Learners learnt a significant percentage of the target items, independently of the experimental condition, making higher progress in word-form learning (20.3%) than in word-meaning learning (9.7%).
2. Having explicit instruction on lexical items prior to the viewing (i.e. through short pre-viewing activities) led to significantly higher gains in both form (+11.8%) and meaning learning (+4.4%).
3. Language of the on-screen text did not have a significant effect on vocabulary gains, in either form nor meaning, but the caption groups tended to perform better than the subtitles groups.
4. Proficiency was a strong predictor of vocabulary gains, for both form and meaning, and particularly for the focused groups. The A2/B1 level group generally outperformed the other two significantly, which may suggest a proficiency threshold at that level.
5. While differences could be observed between learning rates between terms, the effect of the on-screen language, instruction and proficiency was generally consistent across the three terms, independently of other mediating word-related factors.
6. There was a significant, positive relationship between general comprehension and vocabulary gains. This was true for the two captions groups for word-form learning, and for all of them but the SF group. The percentage of variance explained by comprehension was higher for word-meaning and for the groups with access to captions.



7. A considerable percentage of word-forms (M: 64.72%) and word-meanings (M: 61.45%) were retained eight months after the intervention, with no differences between experimental groups.
8. Retention rates for word-meanings were significantly higher when items were presented in written format than when they were presented orally, especially for the two groups in the subtitles condition, suggesting that they might have learnt a higher percentage of word meanings than the first aural test could capture.

## **6.6. Discussion**

### **6.6.1. Vocabulary learning**

Results concerning the first research question showed that, independently of the experimental condition, participants learned L2 vocabulary from extensive exposure to audio-visual input. This concurs with findings from the majority of studies in the field, which claim that L2 vocabulary can be acquired through TV series (e.g., Peters & Webb, 2018; Rogers, 2013; Rodgers & Webb, 2017). Additionally, it was also found that participants made higher progress in recalling form than in recalling meaning – in line with previous studies that showed higher progress in form recognition than in meaning recognition (e.g., Montero-Perez, Peters, Clarebout & Desmet, 2014; Peters et al., 2016). This was found across all conditions, that is, with captions or subtitles, and with or without pre-teaching those words. Importantly, it cannot be forgotten that in the present study recalling meaning is dependent on prior identification of the form, which will have had an effect on meaning recall scores. This setback seemed to be confirmed by comparing results of the two delayed post-tests, as scores in meaning recall were significantly higher when the target items were presented in writing.

The second research question looked into the role that the language of the on-screen text, type of instruction and learners' proficiency had on participants' vocabulary learning regarding form recall and meaning recall. Overall, focused instruction groups performed significantly better than non-focused groups in both form recall and meaning recall, independently of whether they were watching the series with captions or subtitles. This is not surprising, since it is well known from past research that intentional learning is more efficient than incidental learning (e.g., Hulstijn, 2003).

However, although results revealed that language of the on-screen text had no significant effect on either form or meaning learning, it could be observed that in the focused condition, the group with captions outperformed the group with subtitles in both form and meaning recall. This may suggest that when learners are pre-taught the words appearing in the episode they make a first connection between form and meaning of the new words through the pre-viewing activities. Then, having the audio and text in the same language (captions) reinforces the connection between the oral and written form (e.g., Webb & Nation, 2017), which in turn helps recall meaning. On the other hand, when learners are not pre-taught the TIs (more comparable to an incidental learning condition), there is a tendency for the subtitles group to have higher gains in meaning, as they can use the meaning provided by the L1 subtitles to connect it to the L2 oral form, but cannot use this shortcut with captions (and it takes them longer to learn the words). This would also suggest that the L1 text might have compensated for the lack of instruction in the SNF group.

The lack of statistical differences between the captions and subtitles groups falls in line with results from other aforementioned studies (e.g., Bisson et al., 2014; Steward & Pertusa, 2004). If we narrow down the comparison to studies with young viewers, results coincide with those from Bravo's study (2008) – with participants at A2/B1 proficiency level –, in which L1 and L2 groups did not statistically differ, though the subtitles group

performed slightly better. However, Bravo acknowledged that the L1 group was initially more proficient than the L2 group, and since the presence of captions required a higher L2 proficiency level this would explain the lack of differences. Lwo and Lin (2012) also found that varying the language of the text did not have a significant impact on vocabulary gains, and that the effect of different types of text presentation varied depending on learners' proficiency. They found that this was more evident in lower-level learners, who benefited the most from captions, but for advanced learners the presence of the L1 was a distractor – a result that was not found in the present study. On the other hand, the studies by Naghizadeh and Darabi (2015) and by Peters et al. (2016) consistently found that the captions groups performed significantly better than the subtitles groups in vocabulary learning. However, learners in these two studies were older (aged 15-18) than participants in the present study (aged 13), and more proficient. This suggests that there might indeed be an age/proficiency threshold and that the older and more proficient you are, the higher you benefit from captioning rather than subtitling.

In line with the above-mentioned studies, in this study it was found that – as expected – learners' proficiency level was significantly related to vocabulary gains in both form and meaning recall, with more advanced learners obtaining higher gains. (e.g., Chen, Liu & Todd, 2018). Since instruction had a strong effect on vocabulary learning outcomes, the relationship between learners' proficiency and instruction was further investigated, and it was found that the effect of proficiency level (Pre-A, A1 or A2/B1) was different depending on whether TIs were pre-taught or not. Results showed that for form recall participants in the focused groups significantly outperformed the non-focused groups at each proficiency level, and that the A2/B1 group had higher significant gains than A1 and Pre-A. In the non-focused groups, significant differences were only found between A2/B1 and Pre-A levels.

In contrast, for meaning recall the differences between focused and non-focused groups at each proficiency level did not reach significance, although focused groups consistently outperformed non-focused groups. This would suggest that for meaning recall proficiency might have had a slightly stronger effect than instruction. Again, A2/B1 groups in both types of instruction setting significantly outperformed A1 and Pre-A groups. The fact that significant differences were mostly found between the A2/B1 group and the other two suggests the possibility of a threshold at A2/B1 level, over which learners seem to be able to benefit better from exposure to audio-visual input for vocabulary learning. This finding confirms the crucial role played by proficiency and suggests that the results of studies in this area cannot be adequately interpreted if this key variable is not taken into account. In other words, to reach robust conclusions in this line of research, results need to be seen as contingent on the proficiency level of the participants of each particular study. Further research controlling for age or for proficiency can help us conclude which of the two factors has a stronger interaction with the outcomes from either mode.

### **6.6.2. Vocabulary learning and comprehension**

The study has also provided empirical evidence of the relationship between comprehension and vocabulary learning, and confirms the importance of comprehensible input for language learning. There was a positive relationship between general comprehension of the episodes and the overall vocabulary gains, which was stronger in the two captions groups, and also for word-meaning learning. This suggests that having higher comprehension may contribute to learning especially when having access to captions (more demanding than subtitles), and particularly for acquiring meaning, which in the present study was found to be more difficult than acquiring form.

Although understanding the series did not ensure that new words would be learnt, overall, comprehension appeared as a strong predictor of vocabulary learning, independently of the language and the instruction condition. The exception was the SNF group, for which it was found that the relationship between variables – although positive – was not significant. Although learners in the SNF condition were provided with access to word meanings (and had the highest percentage of overall comprehension), they were not provided with the written word forms, either in the on-screen text or in the pre-viewing task.

### **6.6.3. Long-term retention**

Results from the two delayed post-test showed that a considerable proportion of the vocabulary learnt during the last term of the intervention was indeed retained eight months later. The mean retention rate across the groups was of 65% for word form and 61.5% to 74.4% for word meaning (depending on whether retention was assessed through the aural or written input test), with no significant differences between experimental groups. The percentage of retention was significantly higher, however, in the second delayed post-test, which provided learners with the written form of the item instead of the aural version of the word. The aural-format test format requires the listener to identify correctly the word form in order to be able to provide a meaning. Results suggest that participants might have actually acquired more word meanings during the intervention than those that the aural test could capture in the T3's post-test, as they could not identify the word aurally. The considerably large difference between the aural and the written test for the SF group (retention was 25% higher in the written input test) also suggests that, while the L1 facilitated the access to the TIs meaning, having been pre-taught the words did not compensate for not seeing the L2 written form in the captions

during the viewing. The CF group, on the other hand, had on average the same percentage of retention in both tests, which may indicate that access to captions in the L2 had allowed learners to make a stronger aural-written form connection.

#### **6.6.4. Summary**

In sum, Study 2 yields evidence that extensive viewing of TV programmes may support L2 vocabulary learning, and that a considerable proportion of words learnt are retained in the long term. The addition of a small amount of instruction – in the form of a short pre-viewing activity with minimal investment – seemed to make target words more salient and already made a significant difference in learning rates. Arguably, compared to the outcomes of other kinds of vocabulary instruction, the gains in both form and meaning were relatively small (in a total of 515 minutes, participants in the most successful group (CF) learnt on average around 36 word forms and 18 word meanings). However, one needs to bear in mind that vocabulary learning is a gradual, incremental process (Pigada & Schmitt, 2006), and words encountered in the TV series might have been learnt in subsequent encounters, as suggested by the fact that there were also new words learnt between the post- and delayed test. Although it cannot be safely claimed that those words were learnt thanks to the intervention, it is possible that a first contact with them was made during the viewing of the episodes. This would be confirmed by comments made by the students during the follow-up interviews (see Chapter 8), where some of them mentioned that they had indeed encountered again a number of words learnt during the intervention.



## Chapter 7. The role of frequency and imagery

---

### 7.1. Introduction

Research has shown that there are a number of factors that can mediate word learning, such as frequency of encounters (e.g. Waring & Takaki, 2003; Webb, 2014), word relevance (e.g. Peters & Webb, 2018), cognateness (e.g., Puimège & Peters, 2019) and imagery (Rodgers, 2018a; Peters, 2019). While Study 2 has investigated vocabulary gains by comparing different learning conditions, Study 3 will aim at exploring two word-related variables that have been suggested to play an important role in vocabulary learning through viewing (i.e. frequency and imagery) – still underresearched –, while it will also consider other possible mediating factors.

Research on incidental vocabulary learning through reading, listening and, more recently, viewing has provided robust evidence that repeated encounters with unknown words in the input facilitate learning, although the number of occurrences needed for substantial learning remains unclear (Uchihara, et al, 2019). The role of frequency of occurrence has been explored in numerous studies (e.g., Horst et al., 1998; Malone, 2018; Peters & Webb, 2018; Pigada & Schmitt, 2006; Waring & Takaki, 2003; Webb, 2007; Webb & Chang, 2015b), but the number of encounters vary considerably, depending on the mode of input (i.e. reading, listening, viewing) and the conceptualization of word learning (i.e. form recognition, meaning recall) (van Zeeland & Schmitt, 2013). While one encounter might be enough to be able to recognize a word in reading (e.g., Laufer & Rozovski-Roitblat, 2015), many more seem to be needed to recognize meaning, ranging from 2 – 4 encounters (e.g., Pellicer-Sanchez & Schmitt, 2010) to 10 or more (e.g., Pigada & Schmitt, 2006; Waring & Takaki, 2003). It has also



been found that recalling words requires more encounters, especially for meaning recall (e.g., Pellicer-Sanchez & Schmitt, 2010; Webb, 2007). Research has also suggested, however, that the importance of frequency might be less salient in spoken input than in written input (e.g., Brown et al., 2008; Vidal, 2011). Additionally, studies on the effects of repetition in TV viewing argue that frequency may play a different role in this media due to the presence of the images (Peters & Webb, 2018, Rodgers, 2013), but they have generally found that there is a positive effect of frequency on word learning (e.g., Peters et al., 2016; Rodgers, 2013; Peters & Webb, 2018).

Another aspect that has been hypothesised to contribute to word learning is imagery, which is a unique feature of this type of input. A number of studies support the idea that language learners can use the imagery associated with videos to assist information processing. While the benefits of images for language processing have been long acknowledged, research on vocabulary learning through audio-visual input has just started to investigate the extent to which image supports aural information. Building up on the theories of multimedia learning (Mayer, 2001; 2014), Rodgers (2018a) argued that the temporal proximity of the aural utterance of a word and its visual representation may aid word learning, as it facilitates processing by allowing learners to hold separate representation of a word (aural and visual) and build a better mental connection between them. Rodgers found that the images in TV programmes potentially supported vocabulary learning, especially on documentaries, but also in narrative television, even if to a lower degree. Only two studies have looked at the relationship between imagery and gains in this type of input. In a pilot study including university students, Pujadas & Muñoz (2018) found a positive association between co-occurrence of the visual representation of words and the learning rate of those words, and that words that occurred simultaneously visually and aurally were better recalled. Peters (2019), in a study on incidental vocabulary learning through a full-length documentary, also found

that the words that occurred in close proximity to its visual representation were better learnt than the ones without image support.

Study 3 aims at exploring, first, the role of frequency of occurrence in vocabulary gains (form and meaning) in the context of TV viewing, by analysing the effect of repetition of target items occurring in several successive episodes of a TV series. It will also investigate whether the effect of repetition is mediated by other factors, namely the explicit teaching of the target items prior to the viewing, the language of the on-screen text, and the learners' proficiency level. The second focus of the study is to examine the role of imagery in word learning through video, by assessing the relationship between the co-occurrence of word and image and vocabulary gains. It will also explore whether the amount of time the image associated with a target item appears on-screen might mediate learning, an aspect that, to the best of the author's knowledge, has yet to be explored.

## **7.2. Research questions**

Study 3 was designed to address the following research questions:

- Part One
1. To what extent does frequency of encounters predict word-form and word-meaning learning in TV viewing?
  2. Is the effect of frequency mediated by intervention-related variables (i.e. type of instruction, language of the on-screen text) and learners' proficiency?

- Part Two
3. To what extent does the imagery from videos support word-form and word-meaning learning?
  4. Do explicit instruction and learners' proficiency mediate the effect of imagery on word learning?

## **7.3.Part one: the role of frequency**

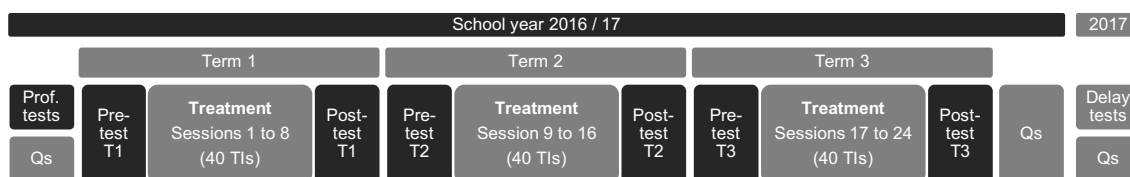
### **7.3.1. Participants**

Participants in the first part of Study 3, concerned with the role of word frequency, were the same as those described in Study 2 (see section 6.3.1). They were 80 students (50 female, 30 male), distributed into four experimental groups: CF (n = 22), SF (n =19), CNF (n = 22), and SNF (n = 17).

### **7.3.2. Materials and procedure**

Because participants were the same as in Study 2, the materials and data collection procedure were also the same. Participants viewed 24 episodes over the course of an academic year, and were exposed to a total 120 target items (TIs), five per episode. For the first part of the study, data from the complete intervention was used (i.e. T1, T2 and T3). The measure of vocabulary learning were the three pairs of pre- and post-tests administered at the beginning and end of each term, and the OPT test scores provided the measure of proficiency. Figure 7.1 highlights the materials used.

Figure 7.1. Materials used in Study 3 (part one)



Word-form and word-meaning learning were scored dichotomously (0 or 1) (see section 6.3.2). Different from Study 2, the analysis was done at the item level, and therefore the overall relative vocabulary gains were not used. Words that were known prior to the intervention were not included in the analysis. The complete list of the 120 target items (TIs) can be found in Chapter 4, and detailed information on word-related features can be consulted in Appendix A.2.

### 7.3.3. Preliminary analysis

The 120 TIs included in the present study were initially selected on the basis that they appeared a minimum of two times in the targeted episode. Frequency of encounters with the TIs within the episode ranged from 2 to 14, with a mean occurrence of 4.46 (SD = 2.5). Because successive episodes were viewed, TIs could reappear in other episodes in the same term. To take into account these additional encounters with the TIs, the total number of occurrences in the eight episodes viewed from pre- to post-test were calculated, and this was used as a (more accurate) measure of frequency in the analysis. Using this measure, items appeared from 2 to 20 times, with a mean occurrence of 5.5 (SD = 3.52). Table 7.1 below (same as Table 4.3 in Chapter 4, repeated here for the sake of clarity) displays the mean frequency of occurrence per episode and per term. There were no significant differences in item frequency between the terms ( $F(2, 117) = .223, p = .801$ ).

Table 7.1. Frequency of occurrence of TIs per episode and term

	n	Frequency within episode				Frequency within term			
		Min.	Max.	Mean	SD	Min.	Max.	Mean	SD
<b>T1</b>	40	2	11	4,02	2,46	2	20	5,45	3,84
<b>T2</b>	40	2	12	4,88	2,67	2	19	5,85	3,70
<b>T3</b>	40	2	14	4,48	2,34	2	15	5,35	3,04
<b>Total</b>	120	2	14	4,46	2,50	2	20	5,55	3,52

As stated above, there were eight sessions between pre- and post-test times, which meant that TIs appearing in the first session of the term were tested up to eight weeks afterwards, whereas TIs appearing in session 8 were tested one week afterwards. Because of the classroom-based and longitudinal nature of the intervention, the time between viewing sessions across experimental groups and between terms was not identical – while the vocabulary testing sessions were. Therefore, although recency of appearance was initially explored and it might have played an effect, it was not included in the final analysis as it was not possible to operationalize this variable consistently.

Table 7.2 contains information on the frequency of occurrence of target items from pre- to post-tests. As can be observed, there was a wide range of encounters: 36.7% of the TIs were encountered between 2 and 3 times, 33.3% between 4 and 6 times, and 30% were encountered 7 times or more. From the 120 TIs, there were 83 (69.2%) that were only encountered in the targeted episode (i.e. massed), while the other 37 (30.8%) appeared in more than one episode of the term (i.e. spaced). Spacing is often strictly operationalized in laboratory studies, in which participants study L2 items in isolation at different time intervals (Uchihara et al, 2019), which was not feasible in the context of

this intervention. Following a similar categorization as the one used in Uchihara et al. (2019), for the present study, TIs occurring only in one episode were considered to be in a massed condition, whereas TIs occurring on more than one episode (and thus, in different days) were in spaced condition. The words that appeared spaced over different episodes are highlighted in **bold** in the table.

Table 7.2. Frequency of encounters with TIs and spacing

	Target items	N (%)
Total frequency of occurrence from pre- to post-test	2 to ace, buckle, cabbage, curly, to gamble, to gloat, haircut, hairdryer, hedgehog, kilt, to loiter, orchids, prank, principal, straight, stuffed animal, threat to tease, to rush, trap to trust	21 (17.5%)
	3 binder, cheers, Christmas Eve, cushy, drill, dump, fake, <b>to fit in, gross,</b> guest, <b>janitor,</b> jukebox, knockoff, <b>to hand out,</b> to hang out, necklace, to mess up, <b>nightmare,</b> oyster, to pick on, teamwork, to trade, whale	23 (19.2%)
	4 to bounce, <b>cranberry,</b> fridge, to ground, <b>to hire,</b> Jewish, nerd, oven, real, rug, shield, sleepover, <b>van</b>	13 (10.8%)
	5 bill, <b>bow,</b> bunk, <b>to cash, to cut off,</b> fortune teller, friendship, to join, <b>proud,</b> to quit, raw, ride, to trip, wealth	14 (11.7%)
	6 bakery, <b>to bet,</b> carpool, chess, to dare, dull, franchise, <b>to figure out,</b> napkin, <b>neighbourhood,</b> rib, to struggle, <b>wall</b>	13 (10.8%)
	7 <b>coach,</b> fireworks, floor mat, <b>to handle, to make sense, realtor,</b> review, <b>scary,</b> turkey	9 (7.5%)
	8 chickenpox, crouton, flight attendant, noodle, <b>ride, to share, tax</b>	7 (5.8%)
	9 <b>grade, hen,</b> piccolo, <b>weird</b>	4 (3.3%)
	10 AC, billboard, <b>costume, mall</b>	4 (3.3%)
	11 to date, <b>deal, fee,</b> to fire, <b>tough</b>	5 (4.2%)
	12 <b>jealous,</b> play	2 (1.7%)
	13 <b>hell</b>	1 (0.8%)
	14 envelope	1 (0.8%)
	15 <b>enough</b>	1 (0.8%)
	19 <b>cool</b>	1 (0.8%)
	20 <b>to move</b>	1 (0.8%)

The mean frequency of occurrence for massed TIs was 4.54 encounters (SD = 2.67), while for the spaced TIs it was 7.81 (SD = 4.15). An ANOVA showed that the differences in frequency between the two types of spacing was significant ( $F(1, 118) = 26.791, p < .001$ ), with spaced items occurring a higher number of times.

### **7.3.4. Results**

#### *7.3.4.1. Word-form learning*

A Generalized Linear Mixed Model (GLMM) with repeated measures was calculated with word-form gains (0 or 1) as the target variable, and frequency of occurrence (total number of encounters), spacing (massed, spaced), type of instruction (focused, non-focused), language of the on-screen text (captions, subtitles) and learners' proficiency (Pre-A, A1, A2/B1) as fixed factors, including all two-way interactions. The model was based on 6892 observations (words known in the pre-test were not included in the analysis).

All variables were first entered in the model and, following a step-back procedure, non-significant interaction and main effects ( $p < .10$ ) were removed one by one until the final fitted model was obtained (see Table 7.3. and Table 7.4). GLMM results revealed that three of the variables contributed to the model significantly: instruction, proficiency and frequency; while the variable spacing emerged as marginally significant. Similar to Study 2, language of the on-screen text did not have an effect on word-form learning. The model did not reveal any significant interactions between the factors.

Table 7.3. GLMM results for word-form learning

Terms	Coeff	SD	t	Sig	Exp Coeff	95% CI for Exp Coeff <sup>a</sup>	
						Lower	Upper
Intercept	-1.643	.2088	-7.867	<.001	.193	.128	.291
Instruction (F)	.771	.1628	4.735	<.001	2.161	1.571	2.974
Instruction (NF)	0 <sup>b</sup>	.	.	.	.	.	.
Proficiency (Pre-A)	-1.056	.2164	-4.878	<.001	.348	.228	.532
Proficiency (A1)	-.640	.1887	-3.394	.001	.527	.364	.763
Proficiency (A2/B1)	0 <sup>b</sup>	.	.	.	.	.	.
Frequency	.051	.0099	5.105	<.001	1.052	1.032	1.073
Spacing (M)	.142	.0740	1.920	.055	1.153	.997	1.332
Spacing (S)	0 <sup>b</sup>	.	.	.	.	.	.

**Instruction:** F = focused, N = non-focused; **Spacing:** M = massed, S = spaced

<sup>a</sup> Confidence interval for Exponential Coefficient

<sup>b</sup> Coefficient is set to zero because is redundant

Table 7.4. GLMM for word-form learning: pairwise contrasts

Factor	Category	$\bar{X}$ (SE)	df	F	p value
Instruction	Focused	.251 (.019)	1, 6886	24.318	<.001
	Non-focused	.134 (.016)			
Proficiency	Pre-A	.122 (.017)	2, 6886	11.013	<.001
	A1	.174 (.017)			
	A2 / B1	.286 (.032)			
Spacing	Massed	.196 (.014)	1, 6886	3.812	.051
	Spaced	.175 (.015)			

Pairwise comparisons showed that differences in proficiency were significant between Pre-A and A1 ( $p = .023$ ), A1 and A2/B1 ( $p = .003$ ) and Pre-A and A2/B1 ( $p > .001$ ). The model revealed that there was a positive relationship between frequency of occurrence and word-form learning. When frequency increased by 1, the odds of a correct response increased by 5.2.%. Spacing contributed to the model marginally ( $p = .051$ ), with massed items receiving around 2% more correct responses than spaced items.



### 7.3.4.2. Word-meaning learning

A second GLMM was run to explore the effects of frequency and spacing on word-meaning learning, following the same procedure as in the previous one. The model was calculated with gains in word-meaning (0 or 1) as the outcome variable, and frequency of occurrence, spacing, instruction, language and proficiency as fixed effects. The final fitted model – based on 7831 observations – is shown in Table 7.5 below. The GLMM revealed that four variables significantly predicted learning of meaning: instruction, proficiency, frequency and spacing, while no interactions emerged. Table 7.6 shows the pairwise contrasts for the significant fixed factors.

Table 7.5. GLMM results for word-meaning learning

Terms	Coeff	SD	t	Sig	Exp Coeff	95% CI for Exp Coeff <sup>a</sup>	
						Lower	Upper
Intercept	-2.545	.2549	-9.981	.000	.079	.048	.129
Instruction (F)	.514	.2034	2.529	.011	1.673	1.123	2.492
Instruction (NF)	0 <sup>b</sup>	.	.	.	.	.	.
Proficiency (Pre-A)	-1.285	.2758	-4.659	.000	.277	.161	.475
Proficiency (A1)	-.740	.2227	-3.323	.001	.477	.308	.738
Proficiency (A2/B1)	0 <sup>b</sup>	.	.	.	.	.	.
Frequency	.060	.0114	5.274	.000	1.062	1.038	1.086
Spacing (M)	.444	.0966	4.599	.000	1.559	1.290	1.884
Spacing (S)	0 <sup>b</sup>	.	.	.	.	.	.

**Instruction:** F = focused, N = non-focused; **Spacing:** M = massed, S = spaced

<sup>a</sup> Confidence interval for Exponential Coefficient

<sup>b</sup> Coefficient is set to zero because is redundant

Table 7.6. GLMM results: effect of categorical variables

Factor	Category	$\bar{X}$ (SE)	df	F	p value
Instruction	Focused	.104 (.012)	1, 7825	6.703	.010
	Non-focused	.065 (.010)			
Proficiency	Pre-A	.047 (.010)	2, 7825	9.639	<.001
	A1	.078 (.011)			
	A2 / B1	.150 (.023)			
Spacing	Massed	.101 (.014)	1, 7825	23.424	< .001
	Spaced	.067 (.015)			

As in previous analysis, the two focused groups significantly outperformed the two non-focused groups, and more proficient students also did better at recalling meaning than less proficient ones. Pairwise comparisons showed that differences were significant between Pre-A and A1 ( $p = .026$ ), A1 and A2/B1 ( $p = .006$ ) and Pre-A and A2/B1 ( $p > .001$ ). The model also showed that frequency was a strong predictor of gains ( $p > .001$ ) and that when frequency increased, the odds of a correct response increased by 6.2%. Finally, in contrast with results for word-form learning, for word-meaning learning spacing emerged as a significant predictor ( $p > .001$ ), with the words encountered in a massed condition being learnt 3.4% more than words that were encountered in a spaced distribution.

#### 7.3.4.3. Exploring other mediating factors

In addition to frequency and spacing of the encounters, there are a wide variety of factors that might also have played a role in word-learning. By way of example, Table 7.7 and Table 7.8 show the 10 word forms and the 10 word meanings that were learnt by the highest percentage of students, alongside several word-related variables. As can

be observed, for word form, frequency of encounters ranged from 2 to 20 (the same range as in the overall sample), and frequency on corpora varied from 1.45 to 215.49 words per million. Unexpectedly, some of the most frequently learnt words forms were only encountered in the first session of the term, which was eight weeks before the vocabulary test. There are also other episode-related factors that might have intervened in making some words more memorable than others, such as the relevance of the word in the episode (e.g., *sleepover*, which was a central topic in the episode were it appeared).

Table 7.7. Word forms learnt by the highest percentage of participants

	Target item	Freq.	Spacing	Recency	SubtLex	Concret.	RG (%)
1	to fire	11	1	7	215.49	2.48	61.11
2	to move	20	2	8	418.14	3.25	55.56
3	to cash	5	2	4	72.43	-	52.94
4	to date	11	1	1	11.73	2.32	51.85
5	to rush	2	1	6	5.25	2.00	50.00
6	sleepover	4	1	1	1.45	3.39	50.00
7	fireworks	7	1	5	5.63	4.86	50.00
8	cool	19	2	8	195.88	3.53	48.84
9	oven	4	1	2	8.88	4.97	47.92
10	to pick on	11	1	8	-	2.48	45.71

**Freq.** = total number of encounters with the TI from pre- to post-test within the term

**Recency:** last session in which the TIs was encountered within the term (from 1 to 8)

**Spacing:** 1 = massed (all encounters within the same episode); 2 = spaced (TI encountered in more than one episode)

**SubtLex** = frequency per million in the SubtLex-US corpus

**Concret.** = concreteness ratings by Brysbaert, Warriner and Kuperman (2014)

**RG** = percentage of relative gains for the item

In the case of word meanings, on the other hand, five of words with the highest percentage of learning were encountered in the last session (i.e. session 8), which indicates that, at least for word-meaning learning, recency might have played an effect – as suspected. Nevertheless, there were, again, frequently learnt words that were only encountered at the beginning of the term (e.g., *turkey*).

Table 7.8. Word meaning learnt by the highest percentage of participants

	Target item	Freq.	Spacing	Recency	SubtLex	Concret.	RG (%)
1	flight attendant	8	1	8	-	4.89	45.21
2	costume	10	2	8	14.14	4.57	38.46
3	whale	3	1	5	11.25	4.96	38.30
4	prank	2	1	8	4.00	2.90	37.10
5	nerd	4	1	8	3.29	3.23	36.76
6	cool	19	2	8	195.88	3.53	31.03
7	hairdryer	2	1	8	.22	4.97	30.99
8	teamwork	3	1	3	2.00	2.37	30.77
9	jealous	12	2	1	38.27	2.17	30.77
10	turkey	7	1	2	22.61	4.89	29.03

**Freq.** = total number of encounters with the TI from pre- to post-test within the term

**Recency:** last session in which the TIs was encountered within the term (from 1 to 8)

**Spacing:** 1 = massed (all encounters within the same episode); 2 = spaced (TI encountered in more than one episode)

**SubtLex** = frequency per million in the SubtLex-US corpus

**Concret.** = concreteness ratings by Brysbaert, Warriner and Kuperman (2014)

**RG** = percentage of relative gains

## 7.4. Part two: imagery analysis

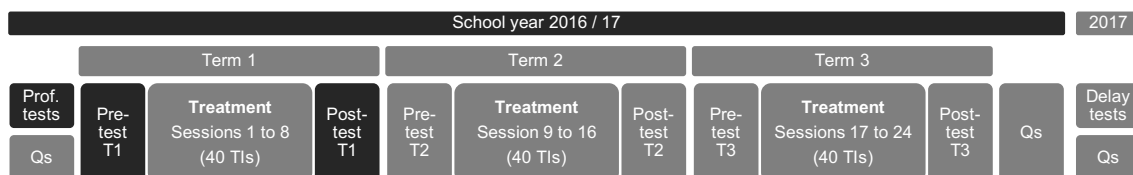
### 7.4.1. Participants

For the second part of Study 3, concerned with imagery analysis, only data from the first term (i.e. T1) was used. Participants with 85% attendance or more during T1 and who had completed T1's pre- and post-test were included in the analysis, leaving a total of 89 students, with an even distribution between the conditions: CF (n = 24), SF (n = 20), CNF (n = 23), and SNF (n = 22)<sup>22</sup>. For part of the analysis, only participants who had also completed the proficiency tests were included.

### 7.4.2. Materials and procedure

For the second part of the study, only data from T1 was used to explore the effects of the imagery support on vocabulary gains. Figure 7.2 shows the instruments used. For this part of the study, only data from one term was used (i.e. T1). This included the first 8 episodes of the TV series and a total of 40 TIs. From those, only nouns (n=20) were selected, due to the nature of the two image-related variables included in the analysis.

Figure 7.2. Materials used in Study 3 (part two)



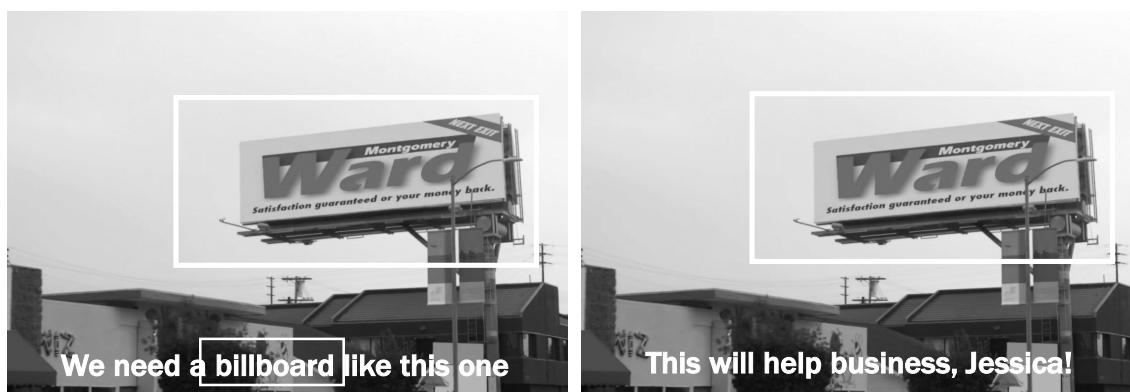
<sup>22</sup> The number of participants in each group is slightly higher because only T1 data is used, and therefore participants that had sessions missing in T2 and T3 could be included in this analysis if they had the required attendance in that term.

In order to investigate the extent to which imagery in the TV series supported aural information, two measures were explored: image co-occurrence and image time on screen. Similar to Rodgers' (2018a) and Peters' (2019) studies, co-occurrence (i.e. CoO) was operationalized as the simultaneous occurrence of the image of a target item and its aural (sound) and written (in the on-screen text) forms in a timeframe of five seconds before or after the occurrence of the item. The rationale behind this timeframe is that the on-screen text only appears for a maximum of six seconds on-screen, and using the +5/-5 seconds limit ensures that the word would have occurred "within an established processing amount" (Rodgers, 2018a: 201). CoO was coded as a binary variable (words were either image-supported or not image-supported).

For the present study a new measure was developed: the image time on screen (i.e. ITOS). ITOS refers to the amount of time the image of a TI is on screen. While research suggests that a higher number of encounters with an unknown word can be conducive to learning, it might be the case that a longer exposure to the image of the word also facilitates learning, and especially word-meaning learning. A longer ITOS – and therefore longer access to the visual, semantic representation of an unknown word – could make the word more salient, and allow L2 learners to have more time to process the information.

Image time on screen was operationalized as the total amount of time (in seconds) in which the visual representation of a TW was present on screen, independently of when or if the TI was uttered simultaneously. The following two images (Figure 7.3) exemplify the differences between the two image-related variables. The image on the left represents co-occurrence of the visual, aural and written form of the TI *billboard* (CoO); the image on the right exemplifies the measure of image on the screen, where the image of a *billboard* is present on the screen (for X number of seconds) without the word *billboard* being uttered (ITOS).

Figure 7.3. Comparison of CoO and ITOS measures



Example of co-occurrence  
of the aural/written word and the image

Example of image time on-screen  
(independently of the aural/written word)

The ELAN software was used to calculate both the CoO and ITOS measures. ELAN is a program developed by The Language Archive (URL: <https://tla.mpi.nl/tools/tla-tools/elan>) that allows to create annotations on video and audio files, and later export them into Excel files for later analysis. Although it is normally used for interview analysis, the fact that this program allows to create annotations frame by frame offered the possibility to determine with high reliability the exact amount of seconds an image appeared on screen. In order to calculate the amount of time each TI's image appeared on screen (i.e. ITOS), an annotation was made for every time the TI image appeared on screen, setting the beginning of the annotation the second the image appeared and closing it the moment it disappeared. Then, the total amount of seconds of each annotation was calculated and each instance was then added up to obtain the total amount of image time on-screen. Figure 7.4 illustrates this process.

Figure 7.4. Sample of the ELAN software output used to calculate ITOS

**TI's time marks**

The screenshot displays the ELAN software interface. On the left, a video frame shows a billboard with a man's face and the text "GREAT FOOD LOW PRICES! Come to the Golden Saddle". An arrow points to the billboard with the label "TI's image". On the right, a table lists time marks for "billboard" with columns for "Nr", "Anotació", "Temps inicial", "Temps final", and "Duració". Below the table is a playback control bar and a timeline. The timeline shows annotations for "billboard" and other terms like "HITS", "buckle", "carpool", and "franchise". Arrows point from the table to the timeline annotations.

Nr	Anotació	Temps inicial	Temps final	Duració
1	billboard	00:01:43.610	00:01:44.899	00:00:01.289
2	billboard	00:01:45.571	00:01:46.882	00:00:01.311
3	billboard	00:01:47.632	00:01:50.343	00:00:02.711
4	billboard	00:02:17.582	00:02:20.725	00:00:03.143
5	billboard	00:08:53.140	00:08:56.517	00:00:03.377
6	billboard	00:09:04.470	00:09:09.244	00:00:04.774
7	billboard	00:09:26.809	00:09:28.470	00:00:01.661
8	billboard	00:12:02.328	00:12:04.505	00:00:02.177
9	billboard	00:12:14.863	00:12:17.821	00:00:02.958
10	billboard	00:12:51.181	00:12:53.895	00:00:02.714
11	billboard	00:19:30.591	00:19:33.250	00:00:02.659

**annotations**

A second rater assessed the TIs' image-support, reaching a 96% agreement. Conflicting cases were discussed until agreement was reached. Note that the location of the TI in the scene (e.g., in a close-up, in the background, being used in the scene) was not taken into account for either of the two measures.



### 7.4.3. Preliminary analysis

For the analysis of imagery, 20 target nouns (TNs) with a variety of characteristics were selected (see Table 7.9). According to the SUBTLEX-US corpus<sup>23</sup>, some were highly frequent (i.e. *ride*), while others were extremely low-frequent (i.e. *real estate*). Concreteness ratings<sup>24</sup> ranged from 2.85 (i.e. *knockoff*) to 4.97 (i.e. *hairdryer*). Frequency of occurrence varied from 2 to 10 encounters, with a mean frequency of 5.25. In most cases, all encounters with the TNs were in the target episode, and only 5 appeared in episodes other than the target (2 had occurrences in prior episodes and 3 in episodes after the one in which they were taught). Distribution of the TNs across episodes was not regular, with one episode having no TNs and one having 5 TNs<sup>25</sup>. Regarding image support, 13 TNs presented co-occurrence of the word utterance and its visual representation, whereas 7 TNs did not. TNs image time on screen ranged from 4 to 128 seconds, with a mean length of approximately 40 seconds on screen.

---

<sup>23</sup> The SUBTLEX-US corpus (available at the Ghent University website, URL: <https://www.ugent.be/pp/experimentelepsychologie/en/research/documents/subtlexus>) is based on American subtitles extracted from 8,388 films and television programmes, and adding up to 51 million words in total (Brysbaert & New, 2009). It has been argued that word frequencies based on film and television subtitles are better predictors of word learning than frequency list based on written sources (Heuven, Mandera, Keuleers & Brysbaert, 2014)

<sup>24</sup> Concreteness ratings according to Brysbaert, Warriner and Kuperman (2014)

<sup>25</sup> The distribution of the 20 TNs across episodes was the following (the amount of TNs per episode appears in brackets): E01 (1), E02 (3), E03 (2), E04 (0), E05 (3), E06 (5), E07 (3), E08 (3).

Table 7.9. Target nouns descriptives

Target Item	Session	Recency	Spacing	Corpus frequency	Concret.	Frequency	CoO	ITOS (seconds)
janitor	1	2	2	5.73	4.68	3	1	4
jukebox	2	2	1	2.27	4.93	3	1	26
napkin	2	2	1	3.61	4.93	6	1	19
crouton	2	2	1	0.25	4.9	8	1	26
nightmare	3	4	2	22.39	2.96	3	0	-
rib	3	3	1	5.9	4.9	6	1	17
mall	5	8	2	18.9	4.83	10	0	-
real estate	5	5	1	0.02	4.25	4	0	-
AC	5	5	1	2.16	4.21	10	1	7
buckle	6	6	1	5.04	4.92	2	1	8
carpool	6	6	1	0.71	3.9	6	1	128
knockoff	6	6	1	0.45	2.85	3	0	-
billboard	6	6	1	1.35	4.83	10	1	29
franchise	6	6	1	2.37	3.72	6	0	-
hedgehog	7	7	1	0.29	4.93	2	1	11
ride	7	7	2	135.37	3.75	8	1	36
principal	7	7	1	13.75	4.79	2	1	91
realtor	8	8	2	1.8	4.61	7	0	-
shield	8	8	1	8.2	4.66	4	0	-
hairdryer	8	8	1	0.22	4.97	2	1	108
<b>Mean</b>	-	-	-	11.54	4.43	5.25	-	39.2

*Session* = target episode where TIs were pre-taught; *Recency* = last episode where TIs were encountered; *Spacing* = massed encounters (1) vs. spaced encounters (2); *Corpus frequency* = frequency per million according to the SUBTLEX-US corpus; *Concret.* = Concreteness ratings by Brysbaert, Warriner and Kuperman (2014); *Internal frequency* = Mean frequency of encounters with TIs within the term; *CoO* = Co-occurrence of TIs and its image; *ITOS* = TIs image time on screen (in seconds).

In sum, the extent to which the image present in the video supports form and meaning learning was explored through two measures. The first one was co-occurrence, a measure used in prior research (Rodgers, 2018a; Peters, 2019). The second one was an experimental measure: image time on screen (the total amount of seconds an image is present on-screen independently of co-occurrence with the aural/written form). Since there is a strong association between them (both variables are measuring the same), analysis were run separately to compare their predictive power.

#### 7.4.4. Results

Data from the T1's pre- and post- test was used to calculate the relative gains in form and meaning for the 20 target nouns (TN). The formula used to calculate them was the same as in Study 2 (see section 6.3.2). Table 7.10 displays relative gains for the 20 TNs appearing in T1 for each experimental group.

Table 7.10. Relative gains for word-form and word-meaning learning in T1

Group	n	Percentage relative gains for TNs (SD)	
		Form	Meaning
CF	24	23.84 (15.85)	12.52 (9.81)
SF	22	18.73 (14.61)	9.18 (8.74)
CNF	23	6.73 (7.95)	3.12 (7.83)
SNF	20	7.45 (8.30)	1.82 (2.54)
Average		14.47 (14.24)	6.86 (8.92)

Welch ANOVAs showed that there was a statistically significant difference between groups for both form ( $F(3,85) = 10.313, p < .001, \omega^2 = .241$ ) and meaning recall ( $F(3,85) = 12.245, p < .001, \omega^2 = .277$ ), and that experimental group accounted approximately for the 24% of the total variance in relative gains in word-form learning, and 27% for relative gains in word-meaning learning. These results for the 20 TNs are the same as in Study 2, where it was found that type of instruction predicted word learning, while language condition did not. Tamhane's T2 post-hoc tests revealed significant differences between the focused groups and the non-focused groups in both form ( $F(1, 87) = 31.127, p < .001, \omega^2 = .257$ ) and meaning ( $F(1, 87) = 25.924, p < .001, \omega^2 = .244$ ), whereas no differences were found within each pair of language groups in either form ( $F(1, 87) = .491, p = .485, \omega^2 = .005$ ) nor meaning ( $F(1, 87) = 1.455, p = .231, \omega^2 = .005$ ). For the following analysis, language conditions will not be further examined.

Table 7.11 shows the number of correct and incorrect responses (and percentage in brackets) for the 20 target items, separated by experimental condition. Items that were known in the pre-test were excluded from the analysis. As can be observed, overall the focused groups had 14.1% more correct responses than the non-focused groups in form, and 8.4% in meaning recall.

Table 7.11. Number (and percentage) of correct and incorrect responses to TNs

	Form		Meaning	
	Correct responses	Incorrect responses	Correct responses	Incorrect responses
<b>CF</b>	101 (23.2%)	334 (76.8%)	56 (12.3%)	399 (87.7%)
<b>SF</b>	75 (18.5%)	330 (81.5%)	38 (9.1%)	379 (90.9%)
<b>CNF</b>	26 (6.3%)	390 (93.8%)	12 (2.8%)	422 (97.2%)
<b>SNF</b>	28 (7.6%)	339 (92.4%)	7 (1.9%)	365 (98.1%)
<b>Total</b>	230 (14.2%)	1392 (85.8%)	113 (6.7%)	1565 (93.3%)

#### 7.4.4.1. Word-form learning

A first GLMM was run to explore the effect of image support on word-form learning. The model was run with form gains (0 or 1) as the dependent variable, and type of instruction (focused, non-focused), proficiency (Pre-A, A1, A2/B1) and co-occurrence (yes, no) as fixed factors. The model was based on 1622 observations. The model revealed – as found in prior analysis – that type of instruction and proficiency emerged as significant predictors (since results for instruction and proficiency were consistently the same across all analyses in the study, they will not be further commented on detail). Co-occurrence did not appear to predict word-form learning, as can be seen in Table 7.12, although words with CoO received a higher percentage of correct responses.

Table 7.12. GLMM results for word-form learning, with CoO

Terms	Coeff	SD	t	Sig	Exp Coeff	95% CI for Exp Coeff <sup>a</sup>	
						Lower	Upper
Intercept	-1.435	.2071	-6.929	.000	.238	.159	.357
Instruction (F)	.947	.1850	5.119	.000	2.578	1.794	3.705
Instruction (NF)	0 <sup>b</sup>	.	.	.	.	.	.
Proficiency (Pre-A)	-1.150	.2361	-4.870	.000	.317	.199	.503
Proficiency (A1)	-.746	.2041	-3.656	.000	.474	.318	.707
Proficiency (A2/B1)	0 <sup>b</sup>	.	.	.	.	.	.
CoO (no)	-.145	.1006	-1.438	.150	.865	.710	1.054
CoO (yes)	0 <sup>b</sup>	.	.	.	.	.	.

**Instruction:** F = focused, N = non-focused

<sup>a</sup> Confidence interval for Exponential Coefficient

<sup>b</sup> Coefficient is set to zero because is redundant

A second GLMM was run again – following the same procedure – with the measure ITOS as the image-support variable. In this model, in contrast with the previous one, ITOS emerged as a significant predictor of word-form learning (see Table 7.13). The model revealed that for every 10 seconds more the image appeared on the screen ( $\exp^{10 \cdot 003} = 1.03$ ), the odds of a correct response increased by 3%, all other factors remaining the same. Therefore, for every minute more in ITOS ( $\exp^{60 \cdot 003} = 1.18$ ), the odds of learning a word form would increase by 18%.

Table 7.13. GLMM results for word-form learning, with ITOS

Terms	Coeff	SD	t	Sig	Exp Coeff	95% CI for Exp Coeff <sup>a</sup>	
						Lower	Upper
Intercept	-1.606	.2086	-7.698	.000	.201	.133	.302
Instruction (F)	.975	.1894	5.150	.000	2.652	1.830	3.845
Instruction (NF)	0 <sup>b</sup>	.	.	.	.	.	.
Proficiency (Pre-A)	-1.224	.2431	-5.035	.000	.294	.183	.474
Proficiency (A1)	-.768	.2087	-3.679	.000	.464	.308	.699
Proficiency (A2/B1)	0 <sup>b</sup>	.	.	.	.	.	.
ITOS	.003	.0004	7.078	.000	1.003	1.002	1.004

**Instruction:** F = focused, N = non-focused

<sup>a</sup> Confidence interval for Exponential Coefficient

<sup>b</sup> Coefficient is set to zero because is redundant

#### 7.4.4.2. Word-meaning learning

The same two models (one with CoO, one with ITOS) were run to assess the effect of imagery on word-meaning learning. The models were based on 1678 observations. In contrast to word-form learning, both imagery variables emerged as predictors for word-meaning recall.

Table 7.14 presents the results from the first GLMM, with CoO as a mediating variable. As it can be observed, CoO was positively related to word-meaning learning ( $F(1, 3105) = 23.289, p < .001$ ), with TNs that occurred simultaneously with their image having higher percentage of meaning learning than TNs without CoO (with CoO:  $M = 6.6, SE = .57$ ; without Co-O:  $M = 2.9, SE = .91$ ). The meaning of words that had the support of imagery were 2.33 times more likely to be learned than words without imagery ( $1/\text{Exp}(\text{Coeff}) = 1/.429 = 2.33$ ).

Table 7.14. GLMM results for word-meaning learning, with CoO

Terms	Coeff	SD	t	Sig	Exp Coeff	95% CI for Exp Coeff <sup>a</sup>	
						Lower	Upper
Intercept	-2.351	.2807	-8.375	.000	.095	.055	.165
Instruction (F)	1.079	.2739	3.939	.000	2.941	1.719	5.032
Instruction (NF)	0 <sup>b</sup>	.	.	.	.	.	.
Proficiency (Pre-A)	-1.608	.3473	-4.631	.000	.200	.101	.396
Proficiency (A1)	-.928	.2595	-3.577	.000	.395	.238	.657
Proficiency (A2/B1)	0 <sup>b</sup>	.	.	.	.	.	.
CoO (no)	-.847	.1693	-5.001	.000	.429	.308	.598
CoO (yes)	0 <sup>b</sup>	.	.	.	.	.	.

**Instruction:** F = focused, N = non-focused  
<sup>a</sup> Confidence interval for Exponential Coefficient  
<sup>b</sup> Coefficient is set to zero because is redundant

Again, another model was run with ITOS as a mediating variable (see Table 7.15). There was also a positive relationship between ITOS and word-meaning learning, which was slightly stronger than for word-form learning: for every 10 seconds more the image

appeared on the screen ( $\exp^{10 \times .004} = 1.04$ ), the odds of a correct response increased by 4%, all other factors remaining the same. If ITOS were to increase by a minute ( $\exp^{60 \times .004} = 1.24$ ), the odds of learning a word meaning would increase by 24%.

Table 7.15. GLMM results for word-meaning learning, with ITOS

Terms	Coeff	SD	t	Sig	Exp Coeff	95% CI for Exp Coeff <sup>a</sup>	
						Lower	Upper
Intercept	-2.802	.2814	-9.958	.000	.061	.035	.105
Instruction (F)	1.059	.2716	3.899	.000	2.883	1.693	4.911
Instruction (NF)	0 <sup>b</sup>	.	.	.	.	.	.
Proficiency (Pre-A)	-1.685	.3511	-4.800	.000	.185	.093	.369
Proficiency (A1)	-.952	.2634	-3.613	.000	.386	.230	.647
Proficiency (A2/B1)	0 <sup>b</sup>	.	.	.	.	.	.
ITOS	.004	.0004	8.038	.000	1.004	1.003	1.004

**Instruction:** F = focused, N = non-focused

<sup>a</sup> Confidence interval for Exponential Coefficient

<sup>b</sup> Coefficient is set to zero because is redundant

## 7.5. Summary of findings

The main findings from Study 3 can be summarized in the following way:

1. Frequency of encounters with unknown target words was positively correlated with learning, with a slightly stronger effect in meaning recall. With increased frequency, the odds of learning a word form or a word meaning increased by 5.2% and 6.2% respectively.
2. Spacing between the encounters was found to be associated with word-meaning learning, with items appearing repeatedly in a single episode (i.e. massed) having higher percentage of correct responses than items with encounters distributed across different sessions (i.e. spaced), even if spaced items had an overall higher number of encounters.

3. The effect of frequency of occurrence and spacing was not mediated by factors related to the intervention (i.e. type of instruction) nor the learners (i.e. proficiency), indicating that the influence of those two variables affected learning rates independently of whether the TIs were taught or not, and independently of the learners' L2 proficiency level.
4. Co-occurrence of the aural/written form of a word with its visual representation predicted word-meaning learning, but not word-form learning.
5. The experimental measure of ITOS emerged as a significant predictor of both form and meaning learning, with a higher effect on word-meaning learning.
6. Co-occurrence and ITOS were not mediated by type of instruction nor proficiency, suggesting that visual support was accessed independency of the intervention-related and learner-related variables in this study.

## **7.6. Discussion**

### **7.6.1. The role of frequency and spacing**

The first part of Study 3 focused on the effects of repetition and spacing on vocabulary learning, including a non-incidental condition, which distinguishes this study from others. Results from the GLMMs showed that frequency of encounters was a significant predictor of both word-form and word-meaning learning, with increasing frequency leading to higher vocabulary gains. This falls in line with results from previous studies on incidental vocabulary learning through audio-visual input, which also found a positive effect of increased frequency on learning (e.g., Peters et al., 2016; Peters & Webb, 2018, Rodgers, 2013). In the present study, half of the participants in the sample had been taught the target items prior to viewing, but the effect of repetition was found to be equal in both focused and non-focused groups. This lack of interaction between instruction and



frequency was unexpected. It would seem that having been pre-taught the words would reduce the need for repetition, since the learners would be aware of the upcoming unknown word and localize it more easily when appearing again. The positive effect of repetition was also found independently of L2 proficiency level.

Specifically, it was found that, all other factors being equal, when frequency increased by 1, the odds of learning a word-form increased by 5.2%, and for word-meaning they increased by 6.2%. This effect was smaller than in other studies in incidental learning (e.g., Peters, et al. 2016). This lower effect may be due to the characteristics of the participants in the sample, who had a lower level of proficiency than in other studies. Another reason might be that there were other aspects that had a stronger effect on learning, such as word relevance (e.g., Peters, 2017; Peters, Noreillie, Heylen, Bulté & Desmet, 2009; Peters & Webb, 2018). The effect of repetition might have been also neutralized by the nature of the study, which could have prompted students to pay more attention to the input in view of the vocabulary post-viewing tasks. In spite of the fundamental difference in the learning setting (e.g., explicit instruction vs. incidental learning) with other studies in viewing, the present study provides evidence that frequency is a potential predictor of learning in this mode of input.

Another aspect that was explored was the role of spacing on vocabulary gains. Results revealed that words that were repeated in the same viewing session were better learnt than words with encounters distributed across different episodes, even though words appearing in a spaced condition had a significantly higher amount of encounters. This could contradict prior research in reading – especially in the case of the focused groups –, which has found that “when vocabulary is learned explicitly, spaced learning is consistently superior to massed learning (e.g., Bahrick, 1984; Janiszewski, Noel, & Sawyer, 2003)” (Webb, 2014: 4). Results would fall in line, however, with the idea that repeated encounters with words in a short span might be more beneficial for learning

(Webb, 2014), which has also been suggested in studies using multiple text or videos to investigate the effect of repetition (e.g., Rodgers & Webb, 2017; Webb & Chang, 2015b; Webb, Newton & Chang, 2012). In the context of television and L2 learning, it is possible that if a word is encountered repetitively within the same episode, it might become more salient to the learners, who in turn may focus their attention on that word and try to guess its meaning from context. On the other hand, if encounters occur in separate episodes, learners might not even notice that there was an unknown word, and if it does not reappear until several episodes later the effect of repetition might be lost.

### **7.6.2. The role of imagery**

The second part of Study 3 focused on exploring the extent to which imagery supports vocabulary learning. The first measure explored was co-occurrence of image and the aural form of the word. Results showed that, while co-occurrence did not predict form learning, it was a strong predictor of meaning learning. The benefits of imagery for word-meaning recall have also been reported by Peters (2019), who found that words that had imagery associated to them were three times more likely to be learned than words without imagery.

The support of the image was also investigated through a new variable: image time on screen. This variable, that takes into account how long the image is present on the screen, emerged as a significant predictor for both form and meaning learning. This would suggest that the longer the visual representation of a target item appears on screen, the higher chances there are that learners will learn that word. For word-form learning, it might be the case that seeing the image of the word makes the word more salient and encourage deeper processing. For meaning recall, a longer appearance on

screen could also facilitate form-meaning connection, as learners do not need to hold the image in their mind but can access it on the screen while processing.

The effect of this variable has been found in both focused and non-focused conditions. This suggests that learners may make use of the image independently of whether they have been pre-taught the target words or not. Considering the focused groups had higher percentage of gains in both form and meaning, it is possible that for them ITOS worked as some kind of reinforcement tool, adding up to the benefits of the instruction received. For the non-focused groups, on the other hand, ITOS might have served as a compensatory mechanism for the lack of instruction. Further research delving into how learners make use of the image in this context –for example, through immediate protocol recalls – could shed light on this matter.

## Chapter 8. The learners' perspective

---

### 8.1. Introduction

Study 1, 2 and 3 have investigated comprehension and vocabulary learning through extensive exposure to a TV series, by measuring several linguistic aspects and the effects of factors related to the input itself, the conditions in which it was presented in the classroom, and learners' prior L2 skills. These studies have confirmed findings from prior research, and have demonstrated that this type of input is, overall, suitable for language learning. They have also provided evidence that students at this age and proficiency level can benefit from audio-visual input. Study 4 seeks to go beyond language benefits – as measured by comprehension and vocabulary tests –, and investigate students' attitudes and perceptions towards audio-visual input as a language learning tool, as well as their feeling of learning from the classroom-based intervention they had participated in. The aim of the study is to offer a qualitative complement to the quantitative findings, and triangulate the data to obtain a wider, richer picture of the learning process.

Past research has suggested that learners' perceptions and attitudes towards audio-visual input as a language-learning tool may affect how learners make use of this type of input. Studies diving into students' perceptions have generally reported that captioned audio-visual materials are viewed positively by EFL learners, since they can help them when the input is beyond their perceived L2 competence. For some students, however, captions can be distracting or confusing, which would hinder their learning potential. It has also been found that the way captions are used is dependent on the learners' L2 proficiency. More advanced learners see them as a backup or a crutch to help confirm what they have heard (Winke et al. 2010), while beginner learners see them as essential

for comprehension (Pujolà, 2002). At the same time, however, low-proficiency students are the ones who report to have more difficulties in benefiting from them. One of the reasons for this might be the lack of familiarity with viewing OV input or with the use of captions. Although past research has generally looked at attitudes at a specific point in time, the few studies that have gathered learners' perceptions after an extended exposure to audio-visual input suggest that attitudes can change over time (Rodgers, 2013; Vanderplank, 2019).

Taking advantage of the longitudinal nature of the present intervention, it was considered that assessing how learners perceive the on-screen text and to what extent they consider it useful for language learning could provide valuable information, which in turn might help interpret learning outcomes. More specifically, Study 4 will examine learners' preferences in relation to on-screen text, frequency of exposure to audio-visual input in the foreign language, and perceived usefulness of TV viewing and on-screen text for language learning purposes. Additionally, it will also explore whether there were any long-term effects from the classroom intervention. This will be done by analysing data gathered through a set of questionnaires administered before and after the 10-month intervention, and data from interviews with a subgroup of students.

## **8.2. Research questions**

There were three main areas of interest in the present study: preferences for on-screen text, changes in frequency of exposure to English audio-visual input, perceived usefulness of audio-visual input as a language learning tool. Study 4 was designed to answer the following research questions:

1. To what extent do learners' preferences towards the use of on-screen text change after an extensive exposure to TV series in the classroom? Are there changes in the language of preference (i.e. subtitles or captions)?
2. What are the main reasons for viewing OV input with subtitles, captions or without on-screen text?
3. To what extent do viewing habits change after an extensive exposure to TV series in the classroom? Are there long-term effects?
4. What is the perceived usefulness of captions and subtitles as a language learning tool *before* and *after* an extended exposure to TV series in the classroom?
5. What is the feeling of learning from exposure to TV series in the classroom? Is it related to the language condition (i.e. subtitles or captions) in which participants had been watching the TV series?

### **8.3. Participants**

As in the prior three studies, the original pool of participants in Study 4 were the 106 secondary school students who participated in the one-year classroom intervention. Data was collected through questionnaires administered at three points in time: prior to the beginning of the intervention, after the intervention and a few months after the end of the intervention. For the first round of questionnaires, participants were the 98 students (60 female, 38 male) who had been regularly attending school during the academic year. For the second round, only participants who had 85% attendance or more during the intervention were included, leaving a total of 88 participants. The number of students in the third round of questionnaires was 86, as two students transferred to a different school after Grade 8.

Additionally, a sub-group of 17 students (12 female, 5 male) was individually interviewed at the end of the second term. Four/five students were selected for each of the four experimental groups, based on their initial proficiency level and their vocabulary gains in T1. This selection aimed at illustrating different viewpoints within each of the groups, while trying to have the same type of students interviewed in each of them (e.g., high-proficient with high gains, high-proficient with low gains, low-proficient with high gain, low-proficient low gains). The same 17 students were interviewed again eight months after the end of the intervention – when the last questionnaire was also administered.

#### **8.4. Materials and procedure**

A detailed description of the four questionnaires can be found in Chapter 4 (see section 4.4.3). The first two questionnaires were given prior to the intervention (September 2016), together with the initial proficiency tests. The first one (BQ1) was given to gather data on preferences concerning on-screen text, prior experience with this type of input, and perception of usefulness of the on-screen text. The second one (BQ2) collected information on participants' out-of-school exposure to English, and served as a measure to determine how representative the group of learners in the study was in relation to a larger population. The third questionnaire was given just after finishing the intervention (June 2017). It assessed possible changes in attitudes and viewing habits, and inquired learners about their feeling of learning from the intervention. A very short, fourth questionnaire (BQ4) was administered eight months after the end of the intervention during the delayed-post-testing phase (February 2018), and was given to assess long-term changes.

Figure 8.1 provides an overview of this timeline in relation to the main intervention. The first round of interviews was conducted by the end of the second term (March 2017), and the second round was conducted during the delayed-post-testing phase (February 2018). All questionnaires were provided in Catalan and Spanish. They were pen-and-paper based, and completed in class under the supervision of the researcher and the teachers. Questionnaires were read in class before filling them in, and students could ask the researcher to clarify any doubt. A sample of the questionnaires can be found in Appendix A.8, A.9, A.10 and A.11, as well as in Appendix B.

Figure 8.1. Materials used in Study 4

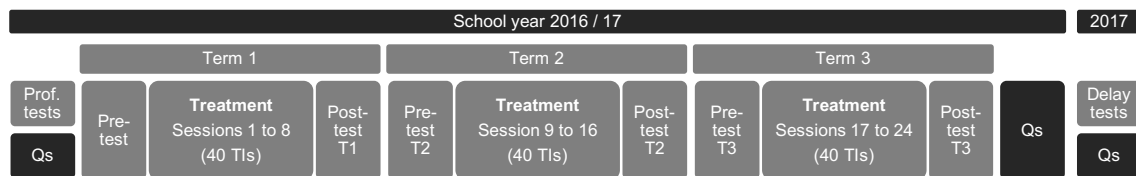


Figure 8.2 illustrates the questionnaires and items compared in Study 4, divided by the three general aspects addressed (i.e. preferences, frequency of exposure, perceived usefulness). The squares in grey indicate the measure used to assess these aspects and in which questionnaires they were assessed. For instance, the preference for or against on-screen text was assessed through a 10-statement measure in BQ1 and BQ3.

Figure 8.2. Assessment of variables across questionnaires

	Preferences for on-screen text			Frequency of exposure		Perceived usefulness		
	10-statement	Self-reported	Why?	Frequency	Changes	5-point scale	Why? (open)	Why? (checklist)
<b>BQ1</b> (09/16)								
<b>BQ2</b> (09/16)								
<b>BQ3</b> (06/17)								
<b>BQ4</b> (02/18)								



## 8.5. Preliminary analysis

In order to assess the extent to which the group of learners in the present sample was representative of a larger population, participants' contact with English outside the classroom was compared to a larger sample of Grade 8 students of the same age and context. One of the background questionnaires administered at the beginning of the intervention (i.e. BQ2) had been previously used in a survey study on out-of-school exposure to English in Catalonia, with data from over 3,000 participants ranging from primary education to university level (Muñoz, forthcoming). Because the items in the questionnaire were the same, this allowed comparing frequency of exposure to English from participants in the current study to data from a subgroup of around 500 Grade 8 students from 15 different schools. Only the first section of the questionnaire – which included items regarding frequency of exposure to English through audio-visual input, gaming, listening to music, reading, speaking and online activities – was analysed.

Mann-Whitney U tests were run to assess whether there were significant differences between the study's sample and the survey's sample. Results showed that, out of the 17 aspects<sup>26</sup> compared between the two groups, differences were only significant in three areas. The first one was frequency of viewing movies and series with subtitles in Catalan or Spanish ( $U = 19557$ ,  $p = .013$ ), with the present study's sample having higher frequency. Secondly, frequency of writing through online platforms ( $U = 19741.5$ ,  $p = .024$ ), with the study's sample having slightly higher frequency. Finally, frequency of

---

<sup>26</sup> The 17 categories analysed were the following: (1) watching movies and series with subtitles in Catalan or Spanish; (2) watching movies and series with subtitles in English; (3) watching movies and series without subtitles; (4) playing videogames in English single player; (5) playing videogames in English multiplayer; (6) playing videogames in English massive multiplayer; (7) listening to music in English; (8) reading books, magazines, comic books in English; (9) talking in English face to face with friends; (10) talking in English face to face with family members; (11) talking in English face to face with tourists; (12) talking in English face to face with abroad; (13) Internet usage: talking online (e.g. Skype); (14) Internet usage: writing in digital support (e.g. email, WhatsApp); (16) Internet usage: watching videos online (e.g. YouTube); (17) Internet usage: listening to music (e.g. podcasts, Spotify).

talking online ( $U = 20600$ ,  $p = .042$ ), with the study's sample having lower frequency. For all the other categories, the samples were comparable. Although generalizations should always be done with caution, overall it seems safe to say that the amount of exposure to English in the present sample of students is fairly representative of what students in Grade 8 would normally be exposed to.

## **8.6. Preferences**

The first two research questions inquired about the extent to which learners' preferences towards the use of on-screen text changed after an extensive exposure to TV series in the classroom, and the main reasons for viewing OV input with subtitles, captions or without on-screen text (i.e. no-text).

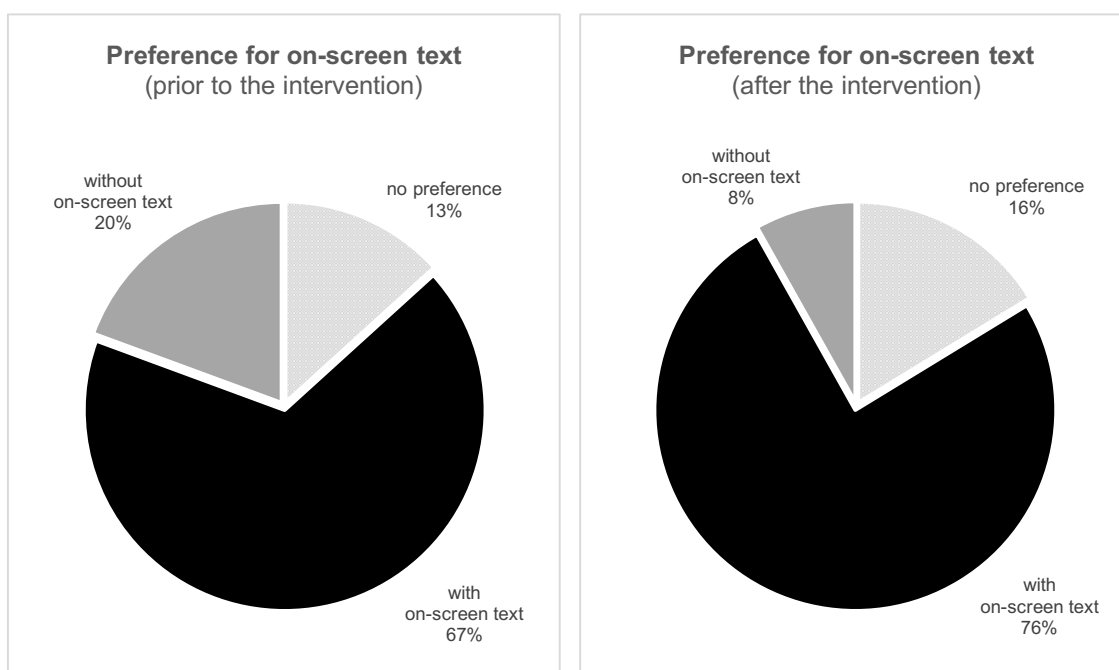
### **8.6.1. Preferences for on-screen text before and after the intervention**

Preference for or against on-screen text was assessed in BQ1 and BQ3 through 10 statements – five in favour of on-screen text and five against it –, and preference was determined by looking at the set of statements with the highest total value (see section 4.4.3.1). Because of its novelty, this measure was piloted with a group of students in Grade 6 and 10 ( $N = 89$ ), and changes were made when a statement was found to be unclear or ambiguous. The Cronbach alpha for the final 10 items was .841.

Figures 8.3 illustrates the difference in preferences for on-screen text before (left) and after (right) the intervention. As can be observed, there was an increase of 9% in the percentage of students who preferred watching OV input with on-screen text (either in the L1 or L2), while the number of students who preferred watching OV input without on-

screen text decreased by 12%. The percentage of participants with no preference for either format remains the same, with a small increase of 3%. It is possible that students who preferred watching input without text support, found captions and subtitles useful after using them for several months. If a learner is unfamiliar with subtitles or captions, they might find them annoying at first, but she / he could become accustomed to them. In Bravo (2008), Grade 8 participants reported that using captions was confusing at first, but that being “forced” to read them made them learn more.

Figure 8.3. Preference for or against on-screen text



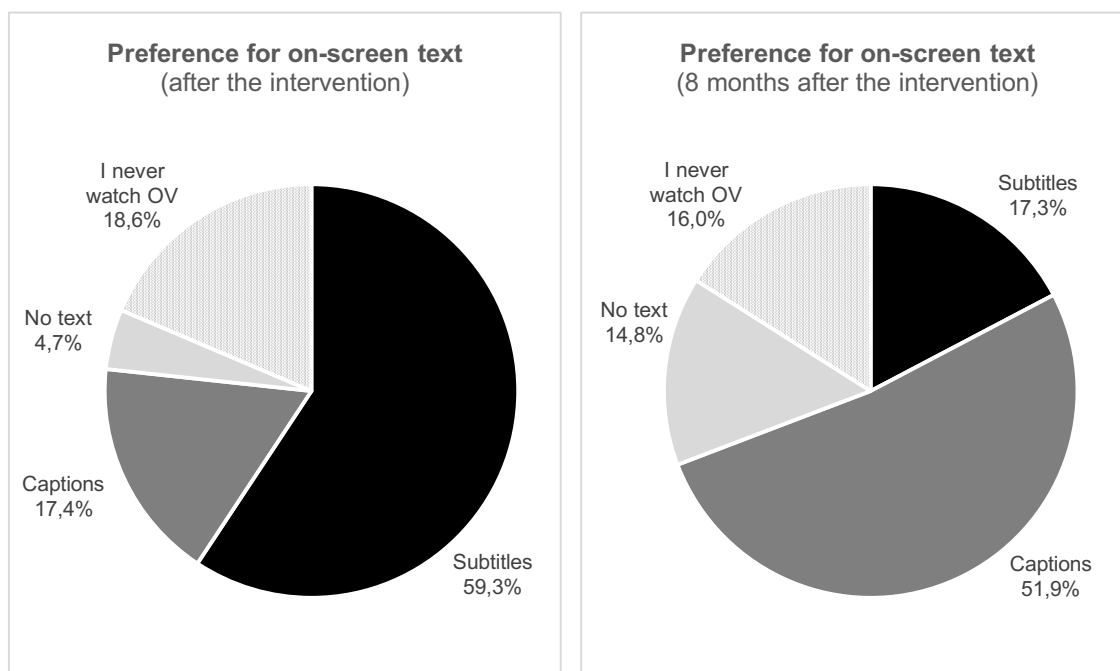
### 8.6.2. Changes in preferences for the language of on-screen text after the intervention

In BQ3 and BQ4 participants were asked directly whether they preferred watching films and series with subtitles in the L1, captions in the L2 or without on-screen text – participants were instructed to selected only their preferred option, even if they also

watched OV input using another format. A comparison was made between the preference provided by the 10-statement measure and this more direct measure, since both items were answered at the same point in time (in BQ3). According to the first measure, 75.6% of respondents were classified as preferring on-screen text, while in the second measure, 76.7% of them checked either the subtitles or captions option. Having answers to the same question using two items in the same questionnaire provides confirmations of its reliability. Therefore, in BQ4 only the second measure – which is shorter and provides more information by differentiating between the languages – was used.

Changes in language preference was measured after the intervention (BQ3) and eight months later (BQ4). As can be observed in Figure 8.4, there was a considerable shift in the language choices: preference for subtitles in Catalan or Spanish decreased by 40%, while preference for captions had increased by 34.5%. There was also an increase in the percentage of people who chose to watch OV input without text support, moving from 4.7% to 14.8%. The percentage of people who reported not watching OV input was similar at both times, with a slight, positive change of the 2.6%. It may be the case that, while it might be possible to train learners to use audio-visual input for language learning, not much can be done when a learner just does not like this type of input.

Figure 8.4. Preference for the language of the on-screen text



### 8.6.3. Reasons behind students' preferences

Participants also reported the reasons for choosing to watch TV series or films with subtitles, captions or without either of them. A list of options was provided for each of the three on-screen text choices (subtitles, captions or no-text), and respondents could choose to answer as many of them as they needed in each language choice (or select the "I never watch OV input" option).

Overall, 47.1% of participants said that they choose to watch films and TV series in OV in English because they prefer to hear the OV over the dubbed version (e.g., the actors' real voices), while 20.9% choose the OV version (with on-screen text or not) because there is no dubbed version available. Fifty percent of the participants would choose to watch the OV *with* subtitles or captions because they need them to understand the content, while 34.9% does it because they consider it helpful. Interestingly, 15.1% of

participants use on-screen text because they watch the series or films with someone (e.g., family, friends) who needs either captions or subtitles.

Table 8.1 reports the most frequent reasons given for each on-screen text option (the percentage of students who choose each option within the total of students who marked that specific on-screen choice). For subtitles in Catalan or Spanish, the main reason given by participants is that they need subtitles to understand the input (56.3%), or that subtitles are helpful although not necessary (37.5%) The third reason is that they do not like captions in English (28.1%). The rationale behind choosing captions in English is similar. The first reason is, again, that captions are needed to understand the video content (37.8%), or that captions are helpful even if not actually needed (27%). The third reason, in contrast, is the lack of subtitles in the students' L1s (24.3%). Finally, the main reason for watching OV without on-screen text is that the on-screen text is not available (50%), which suggest that it is a circumstantial decision rather than a motivated one. A percentage of participants, however, reported that they do not use on-screen text because they find it distracting (25%) and a percentage of students consider that they did not need it (17.9%).

*Table 8.1. Reasons behind the use of on-screen text, by language condition in the intervention*

	<b>Subtitles</b> (N=66)	<b>Captions</b> (N=39)	<b>No-text</b> (N=30)
I like to hear the original version	44.4%	51.4%	64.3%
There is no dubbed version	18.8%	16.2%	21.4%
There is no captioned / subtitled / either	3.1%	24.3%	<b>50.0%</b>
I do not like captions / subtitles / either	28.1%	10.8%	7.1%
I need subtitles / captions to understand	<b>56.3%</b>	<b>37.8%</b>	-
Subtitles / Captions help me understand better (but I do not need them)	<b>37.5%</b>	<b>27.0%</b>	-

I watch series or movies with someone who wants or needs subtitles / captions	14.1%	13.5%	-
I do not need subtitles / captions to understand	-	-	17.9%
Subtitles / Captions distract me	-	-	25.0%
I watch series or movies with someone who does not want / need them	-	-	7.1%

**NOTE:** The percentages in the table are calculated according to the total number of people who answered in each of the columns (i.e. subtitles, captions, no-text). For example, there were 66 students who reported watching OV input with subtitles, and 29 of them (44.4%) said it was because they liked to hear the original version.

The need for on-screen text as a condition to understand the spoken input has been found in prior studies, especially for beginner learners who see them as essential for comprehension (Pujolà, 2002). Although there were no differences between the three proficiency groups in the sample, the majority of participants who reported to need captions or subtitles were generally from A1 level and Pre-A level. Those who said they did not need on-screen text to understand the dialogue were – unsurprisingly – at A2/B1 level, and were also the ones that considered captions and subtitles distracting. There was a weak but significant correlation between no needing on-screen text and finding it distracting ( $r = .289, p = .007$ ), indicating a high correspondence between the participants who selected those two options.

## 8.7. Frequency of exposure

The third research question inquired about the extent to which viewing habits change after an extensive exposure to TV series in the classroom. In BQ1 and BQ4, students were asked to report the frequency of viewing OV input in English (including online videos, such as tutorials or videos from YouTubers) through three multiple-choice

questions, one per each choice of on-screen text – with subtitles in Catalan or Spanish, captions in English or no-text.

Table 8.2 shows the frequency of watching audio-visual input (e.g., films, TV series, YouTube videos) in its original version in English, with either subtitles (in Catalan or Spanish), captions (English), and without on-screen text at two times. The period between the two questionnaires was 18 months.

*Table 8.2. Frequency of watching OV input*

	<b>Frequency of watching English audio-visual input with...</b>					
	... prior to the intervention			... eight months after the intervention		
	Subtitles (%)	Captions (%)	No text (%)	Subtitles (%)	Captions (%)	No text (%)
never	15 (15.3)	36 (36.7)	50 (51.0)	32 (39.5)	24 (29.6)	38 (46.9)
1-2h per year	10 (10.2)	27 (27.6)	15 (15.3)	16 (19.8)	16 (19.8)	11 (13.6)
1-2h per month	15 (15.3)	21 (21.4)	16 (16.3)	11 (13.6)	22 (27.2)	8 (9.9)
1-2h per week	19 (19.4)	8 (8.2)	7 (7.1)	12 (14.8)	5 (6.2)	6 (6.8)
2-6h per week	22 (22.4)	4 (4.1)	5 (5.1)	6 (7.4)	8 (9.9)	9 (11.1)
+6h per week	17 (17.3)	2 (2.0)	5 (5.1)	4 (4.9)	6 (7.4)	9 (11.1)

To make the comparison easier, these six categories were then reorganized into three groups of frequency: “never”, “occasionally” (including 1-2h per year and 1-2h per month) and “weekly” (all the other options). As can be observed in Figure 8.5, prior to the intervention, around 60% of the participants reported watching OV movies or TV series in English with L1 subtitles on a weekly basis, around 10% with L2 captions and 17% without text support. Eight months after the intervention (see Figure 8.6), the percentage of students watching OV input with L1 subtitles weekly went down to 27%, while the



students using L2 captions rose up to 23.5% and students watching OV input without on-screen text support went up to 29%. Data suggest that the 30% reduction in frequency of viewing with subtitles corresponded to participants who shifted their choice to captions and no-text options.

Figure 8.5. Frequency of watching OV input before the intervention

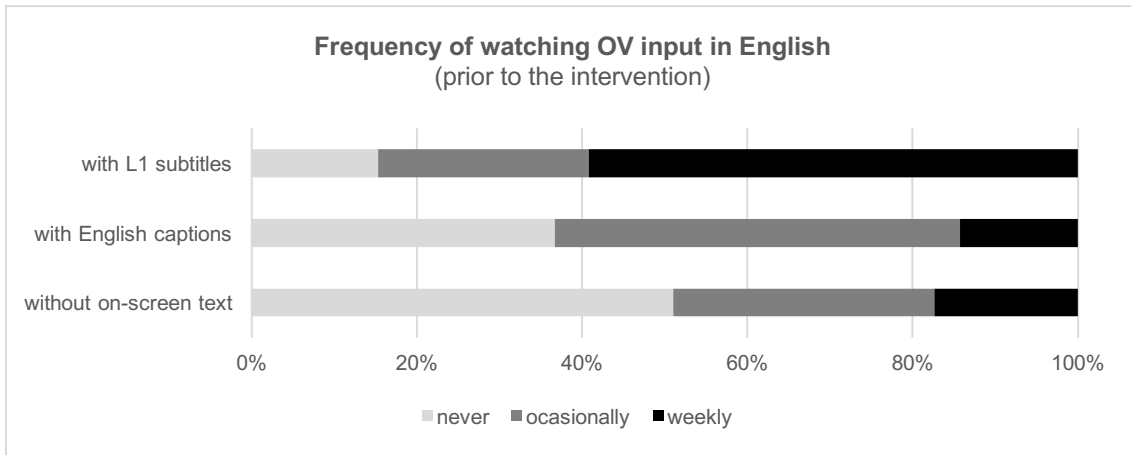
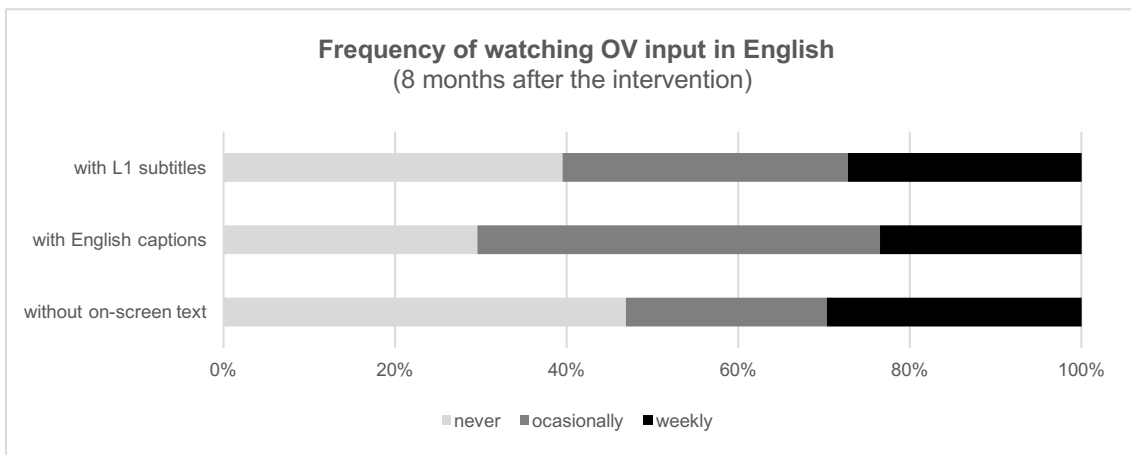


Figure 8.6. Frequency of watching OV input after the intervention



While there was a change in how English input was viewed (with an increase in the use of captions and no-text), there seems to be a similar percentage of participants who did not watch input in English (35% prior to the intervention, 38% afterwards). In BQ4, however, learners were also asked if they had the feeling they were watching more, less or the same amount of audio-visual input in English, and 40.7% of them said they

perceived they watched more input in English. Only 25.9% said that they watched less, and 33.3% said they were viewing more or less the same amount as the previous year. It might be the case that the item format – which asked respondents to calculate viewing in hours from daily to yearly – was not suited for students at this age. A week-based scale might have provided a more accurate measurement.

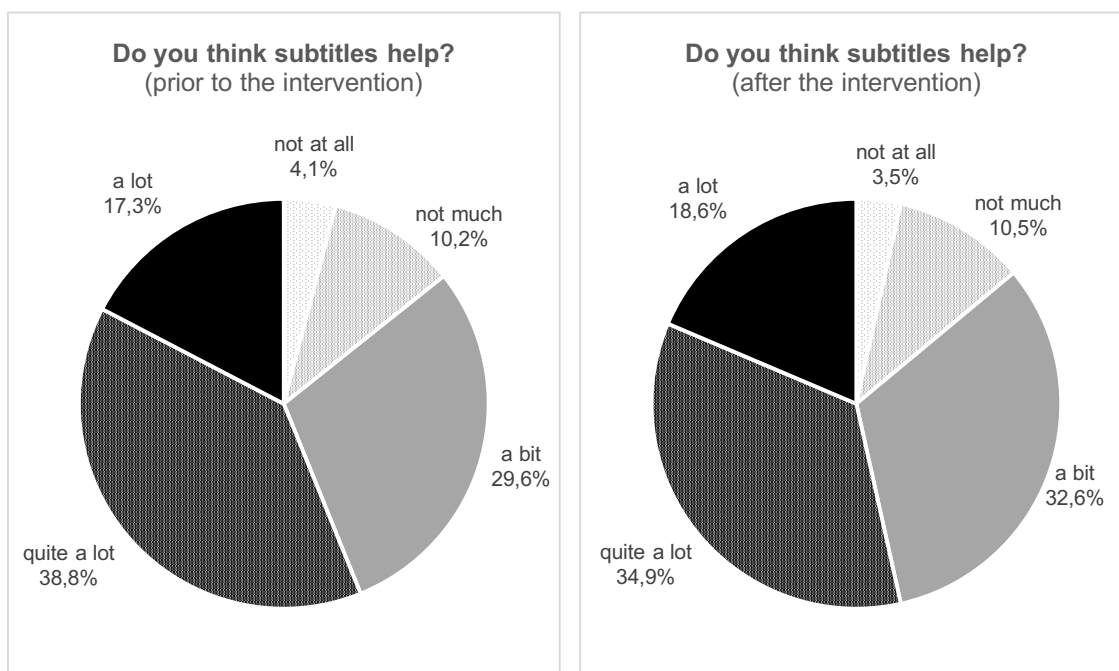
## **8.8. Usefulness**

The last two research questions focused on the perceived usefulness of audio-visual input in English with captions and subtitles as a language learning tool, before and after the intervention, as well as the feeling of learning students had during the intervention.

### **8.8.1. Changes in perceived usefulness**

Prior and after the intervention, learners were also asked whether they thought that having on-screen text was useful for language learning purposes (regardless of the language of the on-screen text) in a 5-point scale from “not at all” to “a lot”. As can be observed in Figure 8.7, prior to the intervention (left) more than 55% of the participants said they found subtitles/captions to be useful or very useful and only 4% considered them useless or annoying. After the intervention (right), their perception of usefulness had remained almost identical, with over half of the participants considering on-screen text to be useful or very useful, and only 3.5% deeming them useless.

Figure 8.7. Perception of on-screen text usefulness before and after the intervention



### 8.8.2. Usefulness for language learning *prior* to the intervention

As a follow-up to the question “*Do you think subtitles help?*”, students were asked to shortly motivate their answer through the open-ended question (“*Why?*”). Comments were coded into 15 different categories, with four major subgroups (learning processes, benefits, drawbacks and perceptions):

#### **Learning processes:**

1. Cross-modal matching (including confirmation of oral / spoken word identification, matching oral and written form or meaning, and identification of (written) form of unknown words).
2. Retention (including facilitation of words’ retention and memorization)

#### **Benefits of the on-screen text:**

3. Content comprehension (plot, dialogues)
4. Listening comprehension (strong accents, speech rate)
5. Pronunciation
6. Vocabulary

7. Contextualization (e.g. understanding cultural aspects)
8. Grammar
9. Writing (spelling)
10. Rising awareness (saliency, paying more attention)

**Drawbacks of on-screen text:**

11. Distracting attention
12. Not having enough time to read Blocks listening
13. Blocking listening

**Perceptions:**

14. Sense of progression (increasing sense of confidence)
15. Proficiency (feeling of improvement)

Each student's comment could fall into more than one category, since students could refer to more than one aspect in their answers. There were 131 coded comments from the 91 students who answered the question. Comments were judged by two raters, with an interrater reliability of 96%. Discrepancies in the classification of the answers were discussed until an agreement was reached, another category was created or the comment was excluded if considered unclear or too vague. The following are illustrations of students' answers (the number in square brackets refers to the code of the comment):

*Perquè diuen les paraules, miro avall i veig el que estan dient i a què es refereix, llavors em quedo amb això* ("Because they say the words, I look at the bottom and I see what they are saying and I know what they are referring to [1], and that is what I remember later [2]" – ID426)

*Perquè així entenc el que diuen i aprenc noves paraules* ("Because like this I understand what they are saying [3] and I learn new words [6]" – ID317)

*Perquè si és una pel·lícula parlen molt ràpid, i entre que llegeixo no escolto* ("Because if it is a movie they speak very fast, and while trying to read [12] I cannot listen [13]" – ID425)

The five highest-ranking categories elicited by students were cross-modal matching (54.95%), general comprehension (20.88%), vocabulary (12.09%), retention (12.09%), and pronunciation (8.79%). A deeper exploration of the data revealed that categories such as grammar or contextualization of cultural aspects were only chosen by the most proficient learners in the sample, while the most proficient students were also the ones that considered on-screen text less helpful.

### **8.8.3. Usefulness for language learning *after* the intervention**

After the intervention, there were two check-all format items regarding the perceived usefulness of the series and the intervention, in which learners had to select – from a list of given options – all of those options they agreed with. The options were based on the most common aspects elicited in the open-ended question from BQ1 and during the individual interviews. The first question (*«In what aspects does the series help you? You can select more than one option»*) focused on aspects related directly with language learning. Table 8.3 shows the percentage of students who selected each category, divided by the language condition they were assigned to in the intervention. For the following analysis, the two groups will be compared<sup>27</sup>.

---

<sup>27</sup> The subtitles group included the SF and SNF groups, while the captions groups included the CF and CNF groups.

Table 8.3. Perceived usefulness for language learning, divided by language condition in the intervention

	Overall N=86	Subtitles n = 41	Captions n = 45
Matching word and pronunciation	39.5%*	26.8%	51.1%
Matching word and meaning	67.4%	65.9%	68.9%
Vocabulary	47.7%**	34.1%	60.0%
Pronunciation	33.7%	29.3%	37.8%
Spelling	33.7%	26.8%	40.0%
Retention of new words or phrases	30.2%**	17.1%	42.2%
General comprehension	25.6%	22.0%	28.9%
Listening in general	46.5%	51.2%	42.2%

\* Differences are significant at the >.05 level

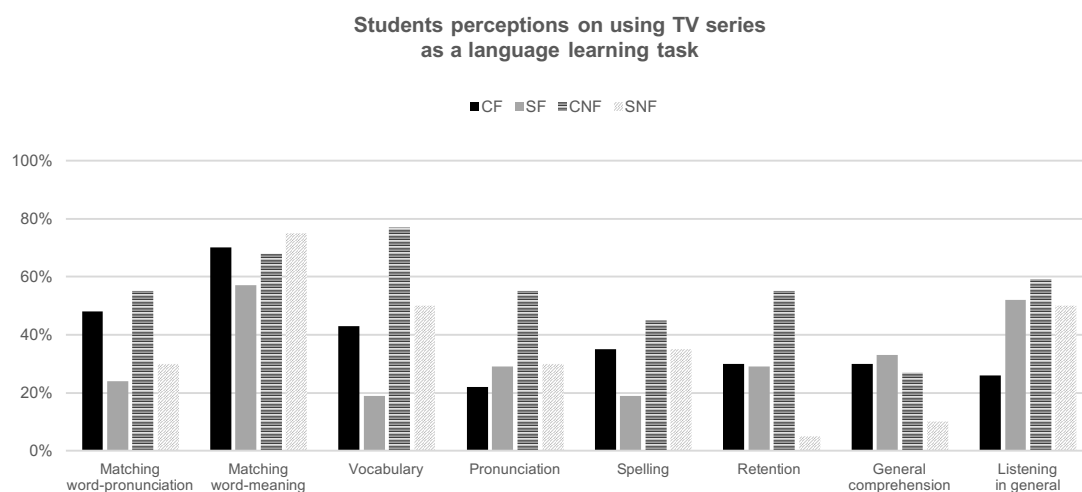
\*\* Differences are significant at the >.01 level

Overall, participants found TV series most useful for making form and meaning connection – regardless of the language of the on-screen text –, followed by vocabulary learning and listening in general. Participants in the captions groups had consistently higher percentage of responses in all categories but one (i.e. listening in general). Results from Welsh ANOVAs showed that differences between the two language groups only reached significance in three cases: matching word and meaning ( $F(1, 84) = 5.569$ ;  $p = .021$ ), vocabulary learning ( $F(1, 84) = 6.035$ ;  $p = .016$ ), and retention of new words and phrases ( $F(1, 82) = 6.963$ ;  $p = .010$ ). In all three cases, it was the captions groups the one with a higher percentage of responses in all three categories.

Differences between focused and non-focuses groups were also explored, but there were no significant differences between the two conditions, except on one aspect: vocabulary. Data revealed that the percentage of learners who considered TV series

useful for vocabulary learning was significantly higher in the non-focused condition (64.3%) than in the focused condition (31.8%) ( $F(1, 84) = 85.518, p = .002$ ). This result was surprising, since focused groups were taught – and had learnt – specific lexical items. It is possible that students who were taught vocabulary explicitly tended to consider the pre-viewing activity – and not the video – as primary source for learning. Figure 8.8 illustrates the differences across the four experimental conditions.

Figure 8.8. Perceived usefulness for language learning by experimental group



It is interesting to compare how perceptions of usefulness varied after the intervention. Table 8.4 below shows the five aspects that students considered that the TV series helped them with in terms of language learning, at the beginning and at the end of the intervention.

Table 8.4. Comparison of perceived usefulness for language learning before and after the intervention

	<b>Most given comments before the intervention (elicited by students)</b>	<b>Most selected options after the intervention (given to students)</b>
1 <sup>st</sup>	Cross-modal matching (54.95%)	Matching word-meaning (67.4%)
2 <sup>nd</sup>	General comprehension (20.88%)	Vocabulary (47.7%)
3 <sup>rd</sup>	Vocabulary (12.09%)	Listening practise (46.5%)
4 <sup>th</sup>	Retention (12.09%)	Matching word-pronunciation (39.5%)
5 <sup>th</sup>	Pronunciation (8.79%)	Pronunciation / Spelling (33.7%)

There was a second item in the BQ3 questionnaire regarding learners' feeling of learning from the intervention, which included eight additional statements. Table 8.5 shows the percentage of students who selected each option, in total and divided by the language group there were assigned to during the intervention.

Table 8.5. Feeling of learning from the intervention

	Overall N=86	Subtitles n = 41	Captions n = 45
"I understand the series better now (e.g., the dialogue, the actors) than at the beginning"	73.3%	75.6%	71.1%
"As a listening activity, I find it more natural than other types of listening"	53.5%	48.8%	57.8%
"I have learnt about cultural aspects (of the Chinese culture)"	52.3%*	63.4%	24.4%
"This type of activity has motivated me"	30.2%	36.6%	37.8%
"I feel relaxed during this type of activity"	52.3%	58.5%	46.7%
"I pay less attention compared to other activities because I feel more relaxed"	18.6%	14.6%	22.2%
"I have the feeling that I have learnt with this activity"	57.0%	46.3%	66.7%
"I would like to continue with this type of activity next year"	81.4%*	90.2%	73.3%

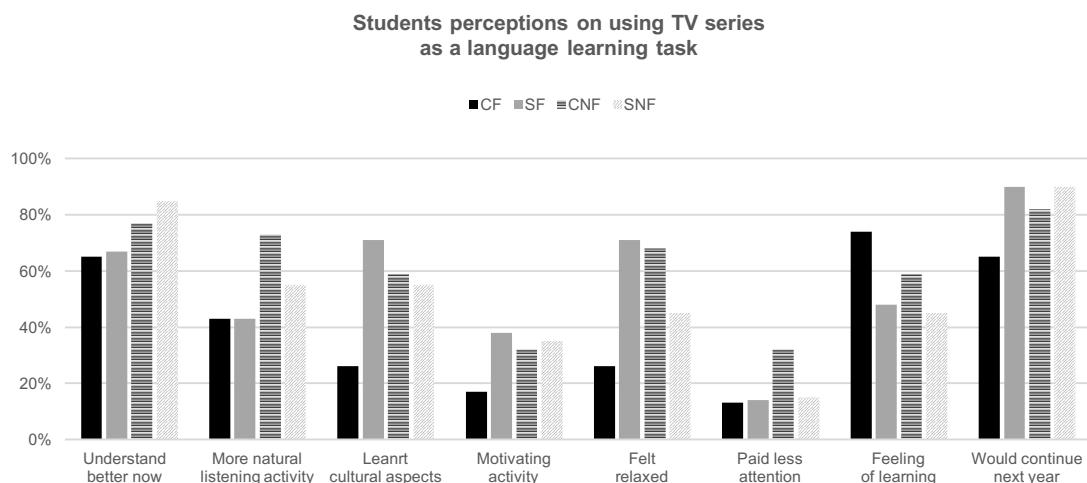
\* Differences are significant at the >.05 level



Compared to the last question, in this case percentages are generally higher for students in the subtitles condition, with the exception of the question directly asking about their feeling of learning. Results from Welch ANOVAs showed that differences between the two language groups only reached significance in two cases: learning about cultural aspects ( $F(1, 84) = 3.959$ ;  $p = .050$ ) and willingness to continue with the activity the following year ( $F(1, 77) = 4.303$ ;  $p = .041$ ), with the subtitles group having a higher number of responses in both cases, but differences are not large in either case. Difference between the groups was marginally significant regarding the feeling of learning, with the captions group being higher than the subtitles group ( $F(1, 82) = 3.667$ ;  $p = .059$ ).

There was a higher percentage of students in the two groups with access to subtitles in the L1 who reported learning about cultural aspects from the series, felt more relaxed during the sessions, found the activity motivating and would choose to continue with it the following year. In contrast, there was a higher percentage of students in the two captions groups who found viewing TV series more natural than other listening activities, and had a higher feeling of learning, although the captions groups had also the higher percentage of students paying less attention. A large percentage of learners (73.3%) reported understanding the series better at the end of the intervention than in the beginning in the same percentage, with no differences between the language groups. Differences across the four experimental condition can be observed in Figure 8.9.

Figure 8.9. Feeling of learning from the intervention by experimental group



#### 8.8.4. Overall assessment of the one-year activity

At the end of the intervention, students assessed the one-year activity in a 7-point Likert-scale, from 0 to 6, and give their opinion about the intervention (see Tables 8.6). The majority of students (88.3%) rated the activity with a 4 or more, and only 7% gave a score below 2. There were no significant differences between experimental groups ( $F(3, 82) = 1.947; p = .128$ ), nor between proficiency levels ( $F(2, 74) = .708, p = .496$ ).

Table 8.6. Overall assessment of the interventions, general and by experimental condition

Assessment of the activity	N	Condition	Mean (SD)
0	4 (4.7%)	CF	4.22 (1.65)
1	0 (0.0%)	SF	4.71 (1.76)
2	2 (2.3%)	CNF	4.73 (1.16)
3	4 (4.7%)	SNF	5.25 (0.72)
4	18 (20.9%)		
5	31 (36.0%)		
6	27 (31.4%)		

Below there are some examples of the students' final comments. Some perceived the activity to have been a listening activity in disguise, and in spite of the vocabulary-focused orientation of the intervention – especially for the focused groups – no reference was made to vocabulary learning. A couple of students also commented they got tired of the TV series chosen after a while, but most of them said they were happy with it, and that they wanted to watch more series the following year, although they asked for a different one. Most of the requests were, however, inadequate for their age, such as Game of Thrones (age 18+) or the Walking Dead (age 16+). This actually points to a problem that many teachers at this age range face when they have to choose audio-materials that motivate adolescents but that are acceptable within the school environment. Some of the comments were the following:

*Està molt bé aquesta activitat divertida camuflada de listening* (“This fun activity, disguised as a listening, it is great” – ID220)

*És molt divertit, ja que no és un listening o un reading normal, sinó que és una història divertida no adaptada* (“It is fun, because it is not a normal listening or reading [activity], but an unadapted fun story” – ID323)

*Odio totes les pel·lícules en general* (“I hate all the films in general” – ID307)

*No aprens res d'anglès perquè només llegeixo i no aprens res* (“You do not learn any English because I only read and I do not learn anything” – ID419)

*Més sèries en anglès, si us plau, estaria bé amb subtítols en anglès i que després parlem de què va* (“More series in English, please, it would be great with subtitles in English, and then to talk about what they are about” – ID423)

## **8.9. Interviews**

Individual interviews were carried out when students were six months into the intervention to take a deeper, and more thorough look at how students were perceiving the intervention as a whole. Students were also interviewed 10 months after the first one to see if changes in attitudes could be observed. Data collected in the first round of interviews were used to improve the end-of-intervention questionnaire (i.e. BQ3), as interviews could rise up issues that would otherwise not have been identified. The main themes discussed in the first interview were (1) comprehension through captions and subtitles, (2) comparison of the intervention with other classroom activities, (3) changes in viewing habits at home, (4) strategies to access multimodal input, (5) perceived learning potential of L2 television, and (6) perception of progress. In the follow-up interviews, students were asked about changes in their viewing habits since the end of the intervention.

### **8.9.1. Comprehension through subtitles and captions**

Students in the subtitles groups were all generally happy with the L1 text, except for one high-achiever, who found subtitles distracting (she reported being used to watching series without subtitles nor captions). When asked whether they could cope with captions, 7 out of 9 said that they consider they could, although all of them preferred subtitles – except the one who preferred no-text. One high-achiever said it could be good to alternate the two languages. Two low-achievers had the perception that viewing with captions would be too difficult, as they considered they would not have enough time to read the text in English and were not proficient enough. One low-achiever aimed at doing so when she achieved higher proficiency. On the other hand, students in the captions groups were overall happy with the L2 text (7 out of 8). The exception was a

low-achiever who felt he would understand the episodes better with the subtitles. Although there were differences in the amount of dialogue they said they could understand depending on their proficiency level, the other seven students reported they preferred captions over subtitles, as they felt that captions forced them to be more attentive and therefore were learning more. Two students considered that subtitles in Spanish were simply useless to learn English.

Overall, most participants were happy – and wanted to continue in – the language condition they had been assigned to, independently of their L2 proficiency level. It is possible that their preference, then, had been affected by the intervention. The majority of interviewees (16/17) thought, independently of the language condition in which they were placed, that it was worth making the effort to use captions.

### **8.9.2. Comparison with other classroom tasks**

The majority of students (15/17) had a higher feeling of learning with the viewing sessions than with other regular classroom tasks. All of them considered the viewing sessions were more fun and enjoyable than regular EFL classes, which according to students were mainly focused on reading, writing, taking notes and copying. Half of the students (9/17) connected higher enjoyment with higher learning, that is, they felt they were learning more because they were having more fun. Two of them were also pleased they could mix something they loved (i.e. watching TV) with language learning. Additionally, four students pointed out that they felt more relaxed during the viewing sessions, and two more recognized that, at some point, they forgot they were in class learning English. While it has been suggested that – due to its leisure-oriented nature –, watching TV series in the classroom could lead to less attention (Vanderplank, 2016),

reports from the students suggest that is not the case. On the contrary, a higher feeling of enjoyment made them generally pay more attention than in other traditional tasks.

### **8.9.3. Impact of the intervention on viewing habits**

At the time interviews were conducted with the 17 students, seven of them reported that they were already watching audio-visual material in English (with or without on-screen text support) at home. One said that she had recently switched from L1 subtitles to L2 captions because of the intervention. Five had begun to watch series and movies in English during the intervention period, motivated by it, while two said they might start to do so. It seemed that the intervention had had an impact on their watching habits at home: from not watching OV to watching with L1 subtitles, from watching with L1 to now using L2 subtitles, from L2 subtitles to no subtitles. Informal conversations in the school also confirmed this, with several students mentioning that they had started to re-watch their favourite TV series in English with subtitles or captions. Many also reported that the fact that they could do so gave them a sense of self-confidence.

### **8.9.4. Strategies to access multimodal input**

Students were also asked about their strategies when viewing subtitled or captioned audio-visual input. In the subtitles group, with access to the L1 translations, 7 out of 9 said they tended to first read the subtitles and then listen to sound, whereas with in the captions groups 5 out 8 stated they tended to first listen to the audio and then read the captions for confirmation or as a crutch for understanding. Proficiency did not seem to have an impact on the strategy followed, but the on-screen text language did: with subtitles there was a general tendency to first read and then listen, and with captions

there is a slight tendency to first listen and then read. A few students in both language groups also mentioned that they tried to make an effort to listen first and only read if necessary.

Students also reported that, when viewing their favourite TV series at home, they would sometimes watch the dubbed episodes first, and then re-watch it again in the original English version, arguing that like this they felt less stressed as they already knew what was going on. During the follow-up interviews, 7 out of 17 students said that they occasionally would stop the video and look up unknown words in the dictionary, especially if they encountered them more than once or perceived they were relevant to understand the content.

#### **8.9.5. Potential for language learning**

Most interviewees (15/17) said that they were learning vocabulary, and the two who did not mention this aspect were in the non-focused group. Also, five students said that the series was useful for improving their listening comprehension. Three high-achievers mentioned they could learn chunks and “how to make sentences”, and another three said that it helped them learn how to structure sentences. Overall, HPs mentioned a wider variety of language aspects – beyond vocabulary learning and general listening comprehension –, and specially in the group with access to captions. On-screen text seems to function as a kind of input enhancement (makes them focus attention on forms), makes students aware of what they understand or know, and what they do not understand or do not know, and allow them to maintain attention in spite of unknown words.

### **8.9.6. Feeling of progress**

All 17 students found that the viewing sessions were easier at the time the interviews took place (six months into the intervention) than at the beginning of the year. They stressed the fact that they had become familiar with the actors' voices, their accents and the characters. For those students who did not have the possibility of practicing oral English in a private school, understanding the series was perceived as a turning point, and there was a great sense of achievement for some of them (e.g., "it boosts my morale to see that I identify a word!"). Overall, there was a strong sense of progression.

### **8.9.7. Long-term effects of the intervention**

Eight months after the end of the intervention, those same 17 students were interviewed again. The majority (12/17) said they were watching more audio-visual input in English now than during the previous year. Also 12 out of 17 learners reported that they used subtitles in Spanish, while only two were using captions in English and three preferred to not use either. Those five students were all high-achievers. The two low-achievers who said that they would start watching TV series in English at home did so, with Spanish subtitles. One of them said she did so when she realized – thanks to the intervention – that she could actually follow them. She considered series helped her to practice English, and while she found captions still too difficult she intended to switch to the L2 text little by little. On the other hand, those students who said they would try to switch to English captions did not, primarily out of laziness – as they admitted. The general use of Spanish subtitles amongst this sub-sample contrast with the on-screen language preference reported by students in BQ4, where around 40% expressed their preference for English captions. It might be the case that, while their intention is to use



captions, they end up switching back to subtitles, and that a bit more practice and higher proficiency are needed before they enjoy watching input with captions.

Students were also asked about what they felt they were learning when watching TV series and movies at home. Learning vocabulary was still the most frequently mentioned aspect (12/17), followed by pronunciation (5/17) and listening practice (4/17). Finally, regarding the words they learnt during the intervention, half of the students (from different experimental groups) said that they did remember some of them, and that they recognized some of them while reading or watching other TV series – a fact that most of them found amusing.

## **8.10. Summary of findings**

The main findings from Study 4 can be summarized as follows:

1. Learners' preferences towards the use of on-screen text increased after the 10-month intervention, and the language preference shifted markedly from Spanish subtitles to English captions in the long term. Data from the interviews suggest, however, that students were still primarily using subtitles.
2. Students reported a wide variety of reasons behind the use of OV input, subtitles and captions. On-screen text – in either language – was primarily used as an essential tool to understand the input, and even when not needed it was considered useful. The original version in English was sometimes chosen just for the lack of a dubbed version, and captions were often used because subtitles were not available, not by choice. Similarly, watching without text support was frequently done because neither captions nor subtitles were available.

3. Viewing habits at home changed after the intervention, moving from subtitles to captions, and from captions to no-text. While it seems that the amount of exposure did not increase, students had the feeling they were watching more input in the foreign language in the months following the intervention. The classroom-based intervention, therefore, seems to have had a positive effect outside the formal setting, suggesting that EFL teachers might need to expose learners to audio-visual input for an extended period of time before students realise they can actually cope with it and decide to use it autonomously.
4. The degree of perceived usefulness of audio-visual materials for language learning purposes increased with an extended exposure. While students could already identify a wide variety of language aspects that could be beneficial for language learning before the intervention, after the intervention the number of learners who could do so was higher, suggesting a rise in awareness.
5. Watching TV series as a classroom activity is perceived as a funnier and more motivating activity than traditional listening exercises, and learners report a high feeling of learning. Overall, perceived usefulness from audio-visual input as a language learning tool was not significantly affected by the language condition, with few exceptions: students who had had access to captions reported higher benefits in vocabulary learning, matching the aural and written form of the words, and word retention than their peers in the subtitles group.



## Chapter 9. General discussion

---

### 9.1. Introduction

This doctoral dissertation has explored the potential of TV series as a language-learning tool through an extensive classroom intervention with adolescent, beginner EFL learners. It has attempted to address gaps in audio-visual research by targeting a younger than usual population, using multiple full-length episodes of the same TV series, and exploring different areas of language learning simultaneously. Through four interrelated studies, this dissertation has examined comprehension and vocabulary learning under different experimental conditions, comparing the effect of the language of the on-screen text and the effect of pre-teaching lexical items. The studies have also taken into account the mediating role of proficiency and learners' prior vocabulary knowledge. Concurrently, learners' attitudes and feeling of learning have been explored to obtain a richer picture of learning beyond the language-based tests. The characteristics of the input used in the study (i.e. the episodes) have also been analysed in order to try to explain differences in comprehension and vocabulary learning rates, by specifically addressing factors that have been shown to be relevant in this context, such as the lexical coverage of the episodes and the potential support that the images provide.

This chapter aims at providing a comprehensive overview of the main findings from the four studies, and triangulate the results across them. It will also highlight the pedagogical implications of the findings, acknowledge limitations, and provide suggestions for further research.

## 9.2. Summary of findings

The main findings from the four studies that compose this dissertation can be summarized briefly as follows:

1. EFL learners at this age and proficiency level are able to benefit from an extensive exposure to TV input with the support of captions and subtitles, attaining an adequate comprehension of the content and acquiring a bigger or smaller part of the targeted lexical items depending on the experimental condition.
2. EFL learners' attitudes are positive and they are motivated to learn language through this type of input, and an extensive, sustained exposure to TV series in the classroom has led to an increased motivation and started a change in viewing habits at home towards L2 TV viewing.
3. For content comprehension, subtitles have been found more efficient than captions, but learners in the caption conditions were able to follow the series too, especially those with higher proficiency level. Vocabulary size emerged, in line with prior studies, as a strong predictor for comprehension when captions were available. When viewing with subtitles, L2 competence was not significant.
4. Comprehension appeared to be episode-dependent, and lexical coverage was found to be a strong predictor of comprehension, independently of other variables. Although no training effect could be observed over the year, the exposure to successive episodes of the same TV series led to higher self-perceived comprehension, as learners reported getting used to the accents and characters over time.

5. Explicit teaching of vocabulary led to higher word-form and word-meaning gains. When words were pre-taught, having access to captions was more beneficial than subtitles. When items were not pre-taught, subtitles seemed to compensate for the lack of instruction, especially for meaning learning.
6. Explicit attention to vocabulary did not affect comprehension significantly, although splitting attention between understanding the content and focusing on lexical items slightly reduced comprehension. At the same time, a higher percentage of comprehension was positively correlated with overall vocabulary gains, specially for word-meaning learning.
7. A high percentage of word forms and word meanings learnt during the last term of the intervention were retained eight months later, with no significant difference between experimental groups.
8. Word-related characteristics predicted word learning. Lexical items that were encountered more often were better learnt, and repetition within the same episode contributed to higher learning than spaced encounters in different episodes, independently of whether the words were pre-taught or not.
9. The imagery in the video supported word learning. The simultaneous presentation of a word and its visual representation led to higher gains in word meaning, while the amount of time an image appeared on screen also predicted the chances of learning a word-form and/or word-meaning.
10. Overall, EFL learners were positive about watching TV series as a language learning tool, and perceived them as more engaging than other traditional listening activities.

### 9.3. Bringing the evidence together

A valuable feature of the four studies that constitute this dissertation is that the participants, the audio—visual materials and the procedure used were the same, which allows for an integration of the data across studies. Table 9.1 summarizes in broad terms the main results across studies, allowing for a comparison between comprehension, vocabulary gains and learners' perceptions across the four experimental conditions. Data from Study 4 (i.e. attitudes and perceptions) appears in italics in the table. While comparison between the measures in the different studies can only be made in broad terms, several outcomes stand out when looking at the results across experimental conditions.

Table 9.1. Cross-sectional comparison of results from Study 1, 2 and 4

	CF	CNF	SF	SNF
Content comprehension	62.77 %	65.73 %	80.23 %	82.99 %
<i>Perceived usefulness for comprehension<sup>a</sup></i>	30.43 %	27.27 %	33.33 %	10 %
Word form gains	30.10 %	13.02 %	21.53 %	14.30 %
Word meaning gains	14.54 %	5.97 %	8.45 %	8.34 %
<i>Perceived usefulness for vocabulary learning<sup>a</sup></i>	43.48 %	77.27 %	19.05 %	50 %
Word form retention	63.6 %	69.6 %	61.9 %	63.8 %
Aural word meaning retention <sup>b</sup>	74.8 %	62 %	56.5 %	52.5 %
Written word meaning retention <sup>c</sup>	73.5 %	76.7 %	81.2 %	66.8 %
<i>Attention<sup>d</sup></i>	1.96 / 3	2.18 / 3	2.57 / 3	2.65 / 3
<i>Enjoyment<sup>d</sup></i>	1.52 / 3	1.77 / 3	2.43 / 3	2.30 / 3
<i>Motivation<sup>a</sup></i>	17.39 %	31.81 %	38.09 %	35 %
<i>Feeling of learning<sup>a</sup></i>	73.91 %	59.09 %	47.61 %	45 %
<i>Overall assessment of the activity</i>	4.22 / 6	4.73 / 6	4.71 / 6	5.25 / 6

<sup>a</sup> Percentage of participants who check that option in BQ3.

<sup>b</sup> Percentage of words retained when the words were presented orally in the delayed post-test.

<sup>c</sup> Percentage of words retained when the words were presented in their written form in the delayed post-test.

<sup>d</sup> Mean value of a variable divided in three categories (low, mid, high), according to their answers in BQ3.

**The results for comprehension and vocabulary go in different directions.** The different combinations of language of the on-screen text and type of instruction affected distinctively comprehension and vocabulary learning. Focused groups learnt a higher percentage of words – thanks to the pre-viewing activities –, while the subtitles groups had higher comprehension – as they had access to the L1. In that regard, then, the SF group enjoyed the most balanced combination, getting the best of both conditions, while the CNF group would be the most challenging one. However, while the SF group did better than the non-focused groups, the CF group outperformed the SF group, suggesting that access to captions was – in combination with explicit instruction, the most beneficial condition for word learning – even though it was the group with the lowest comprehension rate. In contrast, SNF had the highest level of comprehension but limited vocabulary gains, although it did better than CNF. It might be the case that subtitles could compensate – to a certain extent – for the lack of instruction, and especially for word-meaning learning, while instruction could not compensate enough for the lack of captions, especially for word-form learning.

**Learners' perceptions and their actual performance on the language tests do not match.** The percentage of students who perceived that they were learning vocabulary through viewing the TV series was higher in the non-focused groups than in the focused groups, although the non-focused groups were the ones who had the lowest vocabulary gains. It is possible that, in the focused groups, students considered the vocabulary-centred pre-viewing activities more useful for learning than the TV series themselves. On the other hand, a higher percentage of students in the focused groups perceived that viewing the TV series was beneficial for improving general content comprehension, but both CF and SF had a slightly (non-significant) lower comprehension rate than their non-focused counterparts. A reason might be that, learning more vocabulary – through the pre-viewing activities – led to higher feeling of understanding



of the content. This would fall in line with what has been hypothesised in prior research, that pre-learning vocabulary might lead to (or at least lead to perceive) higher comprehension.

**Words learnt are retained in the long term.** A similar percentage of word forms (around 65%) was retained in the long term, regardless of the experimental condition in which the words were learnt. For word-meaning retention, it can be observed that the CF groups (followed by the CNF group) could recall a higher percentage of word meanings when words were presented orally, while the SF group had an advantage in recalling meaning when words were tested in their written form. A possible explanation might be that the CF groups, who had had simultaneous access to the aural and written form of the words, made a stronger connection between the two, and was able to recognize better the words and then translate them in the test. The SF group, on the other hand, might have not developed the same decoding skills, so recognizing the word orally was more challenging. However, since they had access to the meaning, when the words were already provided they could translate them better.

**There is a connection between comprehension, attention, enjoyment and motivation.** Results indicate that the highest the comprehension, the highest the reported attention to and enjoyment from the TV series, as well as the higher the motivation – and vice versa. Granted, while it is not possible to tell the direction nor the exact strength of association between those variables, they seem to go hand in hand.

**Feeling of learning seems to be connected to vocabulary learning, but also to the on-screen language.** In contrast with the relationship pointed out above, the feeling of learning from the TV series seems to be associated with vocabulary learning rather than with enjoyment. The group that reported the highest feeling of learning was the CF (74%), which was the one who learnt overall the highest percentage of vocabulary.

Interestingly, the CNF group – with the lowest gains in vocabulary – reported feeling to learn more than the two groups with access to subtitles. This would suggest that having captions contributed to a higher feeling of learning, even if captions were more demanding and led to lower comprehension.

**Positive assessment of the activity as a whole is linked to affective factors.**

Groups who reported enjoying watching the episodes and were motivated by the activity were the ones giving higher overall scoring to the activity. The SNF was the most positive one, while at the same time being the group with the lowest feeling of learning and some of the lowest vocabulary gains and retention rates. On the other hand, the CF group – who had the higher percentage of vocabulary gains and the higher feeling of learning – was also the one who reported less attention, enjoyed the series the least, reported being less motivated with this activity and had the lowest comprehension rate, all of which probably led them to give the lowest score to the overall intervention.

## **9.4. Conclusions**

This study contributes to the area of foreign language learning through audio-visual input with results from a unique extensive classroom intervention with adolescent, beginner EFL learners. It is the first study analysing learners' exposure to authentic input over an extensive period of 8 months including vocabulary instruction and language of the on-screen text as mediating variables in comprehension and vocabulary gains. It is also one of the few studies that have used several full-length TV programmes (e.g. Gesa, 2019; Rodgers, 2013), and has assessed the long-term effects of an extensive classroom intervention on learners' out-of-school habits. The classroom-based setting in which the study has been carried out yields ecological validity and generalisability to its findings, and complement results more frequently obtained from research with university

students. It is also the first study to investigate the role of imagery in a narrative television programme, and one of the few to assess the effect of repetition across several topic-related episodes.

Results from Study 1 have shown a higher efficiency of L1 subtitles over L2 captions for content comprehension at this level of proficiency, while corroborating the importance of vocabulary size when L2 captions are present. This confirms previous findings in the field, which have also found captions to be more demanding than subtitles for beginner-level learners. The study results also suggest that explicit attention to target vocabulary items may have depleting effects on comprehension scores, which underlines the need to align the cognitive demands of tasks to learners' processing skills. At the same time, adding a pre-viewing task did not significantly hinder comprehension. In view of the benefits brought by explicit vocabulary instruction, including pre-teaching seems to be a valuable language tool. Another relevant finding of this study concerns the interaction between item type and language of the on-screen text, suggesting that learners process textually explicit information and inferential information differently depending on the support they receive from the language available on the screen. The influence of item format and item type on comprehension also highlights the importance of taking into account item-related characteristics when comparing research findings. The analysis of the lexical profile of the episodes revealed (and corroborated) the key role of lexical coverage as a strong predictor of comprehension, in line with findings from prior corpus-driven research and the few experimental studies existing in this area (e.g., Rodgers, 2013).

Study 2 has proven that the integration of explicit instruction and extensive viewing is possible and effective, and results suggests that a small amount of teaching – aiming at directing learners' attention to target vocabulary – is enough to bring about significant improvement, especially on word-form learning. Secondly, it has provided valuable data

for the so-far unresolved issue of the relative gains from captions and subtitles, underscoring the key role played by learners' proficiency and suggesting a proficiency threshold at the A2/B1 level. In that respect, it has been found that the benefits of either on-screen text language depend not only on the linguistic competence itself but also on the instruction condition. The analysis of long-term retention rates has revealed that a high proportion of the word forms and word meanings learnt are subsequently retained. While lexical items might have been encountered again between testing times, it is likely that encountering the words in the episodes may have been a first step in learning them. Overall, the study confirms previous findings regarding the potential of TV programmes for vocabulary learning and as a rich source of comprehensible input.

The exploration of word-related variables in Study 3 provides valuable evidence of the importance of repetition and image support in the area of vocabulary learning through viewing, with data from multiple, successive full-length episodes of the same TV series. Results from Study 3 show that – independently of variables related to intervention (language of the on-screen text, instruction) and the learners (proficiency) –, a higher frequency of encounters with target lexical items contributed significantly to word learning, especially in the case of word-meaning learning. It also revealed that the image associated with target words supported learning in narrative TV, confirming that the simultaneous presentation of a word and its visual representation contributes to learning those words (Rodgers, 2018a; Peters, 2019). An additional contribution is the development and exploration of a new measure of image support (i.e. ITOS), which revealed that image time on screen supported form and meaning learning. While previous research has focused on incidental learning situations, a unique feature of this study is that it concurrently provides evidence from two learning conditions – having explicit instruction and not having it (an incidental-like situation). Findings contribute to the emerging evidence that repeated encounters in topic-related episodes and imagery

support learning, while suggesting that the benefits may not be limited to the incidental learning context.

Finally, Study 4 confirms previous findings regarding language learners' attitudes towards the use of audio-visual materials for language learning purposes, and provides evidence from a population that is typically under-researched. Beginner, adolescent learners – similar to older and more advanced learners – are motivated to learn language through this type of input, which they find more engaging than traditional listening activities, and are willing to use it to improve their L2 skills beyond the classroom. The study also suggests that an extensive exposure within the classroom can have a training effect, modify students viewing habits in the long-term, and endorse TV viewing in the classroom as a starting point for extensive viewing out of the school (Webb, 2015).

## **9.5. Pedagogical implications**

Several relevant pedagogical implications emerge from this classroom-based intervention. Findings confirm the advantage of subtitles over captions for comprehension in a context with limited exposure to English and for adolescent participants with limited proficiency in the target language (average level between A1 to A2) and limited vocabulary size (around 2,000 words). When watching the episode with captions, students with larger vocabulary size could cope better in this more demanding condition, while less proficient students struggled to achieve adequate comprehension. If the goal were to view the series with captions – which have been found to be more efficient for vocabulary learning –, it might be worth contemplating the possibility to do so progressively, by introducing L2 television with subtitles first. In a school classroom setting – where students may have different levels of L2 proficiency – the use of subtitles would engage the weakest students at the beginning while offering all learners the

benefits of listening to authentic input and raising their motivation, moving gradually from subtitles to captions as learners get used to the characters, the voices and the overall topic of the TV series, or alternating between the two types of on-screen text – in a staged approach (Danan, 1992; Montero-Perez, Peters, Clarebout & Desmet, 2014). Even if full comprehension may not be achieved at the beginning, EFL teachers (and learners) should bear in mind that the goal is for comprehension to increase over time (Webb, 2015).

The association of the episodes' comprehension and lexical coverage suggests the need to align the vocabulary load of the audio-visual input to learners' language skills. Language teachers should be very careful when selecting the viewing materials for the classroom, and provide additional support if the input is beyond students' L2 skills, especially when there are learners with a range of proficiency levels. A way to do so would be, as it was done in the present intervention, to add captions or subtitles, or pre-teaching frequently occurring words in the episode. Increasing vocabulary is – ultimately – an efficient way of supporting comprehension of L2 input. In the study, a minimal time investment on pre-teaching with receptive activities helped students' learning of vocabulary, which could be further enhanced with additional productive activities (Chung, 1996; Sockett, 2014). Such activities – which in this case consisted on short 5-minute tasks – could be easily created by EFL teachers, and embedded in the regular classes so students focus on language learning and do not see the viewing activity as just entertainment (Webb, 2015).

The benefits brought by word repetition and the presence of word-associated imagery also highlights the importance of adequately pre-selecting the target lexical items. Targeting words with a high number of encounters within the same episode, and (or) words that are image-supported, could further promote learning. While carrying out a detailed, in-depth study of the episode's vocabulary and imagery would probably place

an excessive burden for EFL teachers, a quick analysis of the lexical profile through tools such as Lexical Tutor would allow teachers to roughly assess the overall vocabulary load of the video input and reveal which words re-occur more often. A useful recommendation would be that EFL teachers become aware of the language learning potential of this type of materials, and that they are trained on relevant profiling strategies. This would be useful not only to pre-select programmes in class, but also to suggest adequate-level TV programmes and movies for students to watch at home. Related to this, the creation of a database or online platform by language experts including TV series' lexical profiles and/or potential unknown vocabulary lists specific to those TV programmes would be a valuable resource for EFL teachers and students alike.

Data from the questionnaires and interviews also revealed that learners were very positive about and appreciative of encountering 'real' English in class, which is of special interest for teachers. Students were paying attention to the input and had a strong feeling of learning, and the leisure-oriented nature of the TV series did not seem to distract them (cf. Vanderplank, 2016). Learners reported that they were even more attentive, as they found the TV series more entertaining and engaging than regular class activities. More importantly, after some extended exposure in the classroom, some students started changing their viewing habits at home, and began to watch (o re-watch) series and movies in English as they felt they could understand them better. Especially in settings where L2 input is limited, teachers may introduce L2 television viewing in the classroom to promote autonomous extensive viewing outside the classroom, and they may train students to enhance their learning potential with strategies and focus on form. As summarized by Webb (2015):

"The primary aim of extensive viewing is to encourage regular independent out-of-class L2 television viewing after initial classroom-based viewing. The purpose of a classroom-based viewing program is to raise awareness of the benefits of L2 television for language learning

teach learners strategies that can be used to support their comprehension, and demonstrate that through implementing a principled extensive viewing approach, comprehension may be sufficient for pleasurable viewing” (Webb, 2015: 2)

An interesting finding from the study is that participants in the captions groups recognized that captions were more demanding than subtitles, but considered that it was worth making the effort – independently of their L2 proficiency level. On the other hand, participants in the subtitles condition did not want to switch to captions due to their self-perceived limited L2 competence. This would suggest that students eventually became used to the language condition to which they had been assigned, and that if teachers were to “force” students to use captions in the classroom, they would eventually get used to them.

Finally, the study also had societal implications demonstrating to students and EFL teachers the value of watching captioned/subtitled TV series as an L2 resource in a traditionally dubbing country. Hopefully, the study has contributed to a change of viewing habits in the classroom and at home, and has increased learners’ exposure to a much needed naturalistic, comprehensible input in the foreign language.

## **9.6. Limitations and future research**

This doctoral dissertation has some limitations that need to be acknowledged. First, due to its longitudinal nature, although there was a considerable sample of participants, the final number of participants in each experimental group was relatively small, especially when proficiency levels were compared within group. At the same time, however, the fact that comprehension and vocabulary gains were evaluated through a considerable number of tests increased the results reliability. Because the study was



conducted during regular EFL classes – the viewing conditions were dependent on the school setting, and levels of attention probably cannot be comparable to that of a university classroom or a language laboratory. Related to this, another shortcoming is that the studies did not take into account the degree of attention participants were actually putting into the tasks, which could be a concern when extrapolating results. Although more faithful to the real learning environment, this is a common shortcoming of classroom studies, which may be seen as an inevitable concomitant of their ecological validity. Similarly, this environment precluded the existence of a control group as ethically, their learning opportunities without any exposure to the TV series in English might have been reduced.

There are also a number of limitations relative to each of the studies. In Study 1, the number of items per test was relatively low for a 20-minute video, especially compared to other studies in the field – with a much higher ratio of item to video-minutes, although there was a considerable number of observations after adding up the 24 tests. This was done so the items in the test did not overlap, and that information present in a question or the distractors did not hint the correct answer in another – especially considering that detailed questions were avoided and there was a limited amount of information to be asked for in a single episode. A second reason was to keep the test short and avoid overwhelming students – as students also had to complete vocabulary tasks within the viewing session.

Another shortcoming from this study was that attention to and enjoyment from the TV series were only assessed at the end of the intervention. Further research including these questions at the end of each session (e.g., Rodgers, 2013) might provide a more accurate measure of these variables and allow comparing their perceptions to the actual level of comprehension of the episodes. Familiarity with viewing OV input was calculated based on a combination of measures and provided a general measure of frequency of

use, but two separate measures – one for captions, one for subtitles – might have been more accurate. Neither did the study take into account a number of factors that might also have an influence on comprehension. Such factors could be the location of the necessary information within the episode, since items tend to be easier when the information is presented at the beginning or when it is repeated (Buck, 2001). Another factor would be whether the answer to a question included a lexical item taught in the focused groups, in which case the focused groups could have had an advantage. Topic complexity of the episodes (e.g., cultural-bound reference) could also be explored. Further research including these variables, other TV genres, and other proficiency levels would provide valuable information and help obtain a more comprehensive picture of the factors involved in TV viewing comprehension.

Study 2 also presents several limitations. First, the type of test used to evaluate learning might explain the low gains obtained – especially in meaning – since a recall test (e.g., a translation test) is more difficult than a recognition test (e.g., multiple-choice test) (Jones, 2004). If a student failed to identify orally a target item in the first place, they could not provide a translation, but this does not mean the participants could not recognise the word form if encountered, or that they did not know the meaning of the word. A second test to check partial knowledge of the target items may have provided a more accurate picture of their learning, as was done in the delayed-post-testing phase. The difference between meaning recall in the two types of test (i.e. aural vs. written) indicates that the aural test used in the intervention might have failed to measure partial knowledge of the word meanings. As regards form, the requirement that words had to be correctly spelt might have put the subtitles groups in disadvantage. There could also be other factors playing a role in being able to transcribe word forms, such as the word's phonological distance to the learners' L1 (i.e. phonological opacity).

Regarding the analysis of the intervention's long-term effects, it should be noted that a considerable amount of time passed from the T3's post-test to the delayed post-test. On the one hand, this provides valuable evidence of long-term retention, as studies have generally tested retention through tests administered from a week to two months after the post-test. On the other hand, as commented before, it is also likely that learners were exposed to the TIs during the 8-month gap, and that some of the words were learnt during that period – as indicated by the small percentage of new words correctly recalled in the delayed test. In any case, the viewing of the episode provided the first step in these words' learning.

Study 3 was concerned with word-related variables and it primarily focused on two aspects: frequency of encounters with the target items and the analysis of the imagery associated to them. However, the study did not take into account other word-related factors, such as part of speech, concreteness or frequency of occurrence in corpus (e.g., SUBTLEX corpus, BNC/COCA word lists), which have been shown to play a role in word learning. While it would have been interesting to explore those aspects, it was considered that the reduced number of target items and the variability within the sample – in addition to the four learning conditions and the diversity of proficiency levels – would hinder the reliability of the results. Another shortcoming of the study is that the effect of recency was not examined, although it is possible that words appearing in viewing sessions closer to the post-tests (i.e. session 7 or 8) had had higher rates of learning.

Because the viewing materials were authentic, full-length episodes of a TV series, there was also little control over the items distribution across the sessions nor the number of repetitions per item. Target items had a frequency of occurrence ranging from 2 to 20, which has been shown to be a positive feature when assessing the effect of repetition (e.g. Uchihara et al., 2019), but the variability in range was large: around a third of the items occurred between 2 and 3 times, while another third occurred between the

spectrum of 7 and 20. The number of massed and spaced items was also unbalanced, and a similar drawback was found for image-supported items. In the case of imagery, moreover, only a small number of items could be analysed. Findings from this exploratory analysis, however, provided initial evidence that the image associated with videos support word learning.

Finally, regarding Study 4, a main shortcoming of the questionnaires was that variables such as attention and enjoyment were only measured at the end, although it is likely that these two variables fluctuated from episode to episode. In both BQ1 and BQ3, items regarding the usefulness of on-screen text and audio-visual materials did not specifically distinguish between captions and subtitles. Though it was assumed that learners answered the questions with their language condition in mind, a separate evaluation of the benefits of either language in the on-screen text would have been more accurate. The fact that questionnaires were not anonymous could also have prompted students to tend towards answers that are more positive. However, the rapport built between the researcher and the students suggests that they were familiar enough to answer truthfully, which was confirmed by the student-researcher interactions throughout the intervention and during the interviews.

## **9.7. Final remarks**

This doctoral dissertation has presented a comprehensive overview of language learning through extensive TV viewing, and has provided evidence of the benefits of this type of input for comprehension and vocabulary learning. The study contributes to the growing area of research on viewing with data from a longitudinal classroom-based study with adolescent, beginner EFL learners. This doctoral dissertation has tried to bring together – by comparing the results from the four studies that constitutes it – results from

different aspects of language learning, and it has made several original contributions to the field.

The study has used 24 full-length successive episodes from the same TV series (515 minutes), allowing for the accumulation of background knowledge and the vocabulary repetition that has been suggested to benefit comprehension and incidental vocabulary learning. It has also compared two different instructional conditions, and has found that pre-teaching lexical items significantly contribute to vocabulary learning, providing evidence to confirm prior findings in corpus research (e.g., Webb, 2010b). The study has also explored the improvement in comprehension and vocabulary gains over time, and has also investigated the relationship between these two variables. It is one of the first studies to explore the role of imagery and its direct association with vocabulary gains, and has introduced a new measure to evaluate the imagery support. Finally, the study has also explored students' perceptions and attitudes, presenting data from a non-university population, and more importantly, it has assessed how perceptions changed over time. In line with the principles of extensive viewing (Webb, 2015), these findings also support the use of extensive viewing programmes in the classroom – which would be already appropriate at this age and proficiency level – as a starting point for successful, autonomous L2 television viewing at home.

## References

---

- Adolphs, S., & Schmitt, N. (2003). Lexical coverage of spoken discourse. *Applied linguistics*, 24(4), 425-438.
- Almeida, P., & Costa, P. (2014). Foreign language acquisition: the role of subtitling. *Procedia-Social and Behavioral Sciences*, 141, 1234-38.
- Aloqaili, G. S. (2014). *Learning Vocabulary from Subtitled Videos: An investigation into the effectiveness of using subtitled videos for intentional vocabulary learning in Saudi Arabia with an exploration of learners' perspective*. (Unpublished master's dissertation). University of Southampton.
- Alptekin, C., & Erçetin, G. (2010). The role of L1 and L2 working memory in literal and inferential comprehension in L2 reading. *Journal of Research in Reading*, 33(2), 206-19.
- Bairstow, D., & Lavour, J.M. (2011). Audiovisual information processing by monolinguals and bilinguals: effects of intralingual and interlingual subtitles. *Approaches to Translation Studies*, 36, 273-93.
- Baltova, I. (1994). The impact of video on the comprehension skills of core French students. *Canadian Modern Language Review*, 50(3), 507-31.
- Baltova, I. (1999). *The effect of subtitled and staged video input on the learning and retention of content and vocabulary*. (Unpublished doctoral dissertation). Ontario Institute for Studies in Education at the University of Toronto, Ontario, Canada.
- Bavaharji, M., Alavi, Z., & Letchumanan, K. (2014). Captioned instructional video: effects on content comprehension, vocabulary acquisition and language proficiency. *English Language Teaching*, 7(5), 1-16. doi:10.5539/elt.v7n5p1.
- Bianchi, F., & Ciabattini, T. (2008). Captions and subtitles in EFL learning: an investigative study in a comprehensive computer environment. In Baldry, A., Pavesi, M., & Taylor Torsello, C. (Eds.) *From didactas to ecolingua* (pp. 69-90). Trieste: Edizioni Università di Trieste.
- 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.

- Birulés-Muntané, J., & Soto-Faraco, S. (2016). Watching subtitled films can help learning foreign languages. *PloS One*, *11*(6): e0158409. doi:10.1371/journal.pone.0158409.
- Bisson, M., Van Heuven, W., Conklin, K., & Tunney, J. (2014). The role of repeated exposure to multimodal input in incidental acquisition of foreign language vocabulary. *Language Learning*, *64*(4), 855-77. doi:10.1111/lang.12085.
- Boers, F. (2018). Intentional versus incidental learning. *The TESOL Encyclopedia of English Language Teaching*, 1-6.
- Bonk, W. J. (2000). Second language lexical knowledge and listening comprehension. *International journal of listening*, *14*(1), 14-31.
- 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.
- Bravo, M.C.C. (2008). *Putting the reader in the picture. Screen translation and foreign-language learning* (Unpublished doctoral dissertation). University Rovira i Virgili and University of Algarve.
- Brett, P. (1996). Using multimedia: An investigation of learners' attitudes. *Computer Assisted Language Learning*, *9*(2-3), 191-212.
- Brett, P. (1998). Using multimedia: A descriptive investigation of incidental language learning. *Computer Assisted Language Learning*, *11*(2), 179-200.
- 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.
- Bruton, A., García López, M. & Esquiliche Mesa, R. (2011). Incidental L2 Vocabulary Learning: An Impracticable Term? *TESOL Quarterly* *45*(4): 759-68. 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*(4), 977-990.
- 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.
- Buck, G. (2001). *Assessing listening*. Cambridge: Cambridge University Press.

- Caraker, R. (2016). Spain and the Context of English Language Education. *研究紀要* [Research Summary], 92, 23-35.
- Chai, J., & Erlam, R. (2008). The effect and the influence of the use of video and captions on second language learning. *New Zealand Studies in Applied Linguistics*, 14(2), 25.
- Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and instruction*, 8(4), 293-332.
- Chang, A., & Read, J. (2006). The effects of listening support on the listening performance of EFL learners. *TESOL Quarterly*, 40(2), 375-97.
- Chapple, L. & Curtis, A. (2000). Content-based instruction in Hong Kong: Student responses to film. *System*, 28, 419–33.
- Charles, T., & Trenkic, D. (2015). Speech segmentation in a second language: The role of bimodal input. In Gambier, Y. Caimi, A., & Mariotti, C. (Eds.) *Subtitles and language learning: Principles, strategies and practical experiences* (pp. 173-198). Bern: Peter Lang
- Chen, Y., Liu, Y. & Todd, A.G. (2018). Transient but Effective? Captioning and Adolescent EFL Learners' Spoken Vocabulary Acquisition. *English Teaching & Learning* 42(1), 25-56. doi:10.1007/s42321-018-0002-8
- Cheng, H. (2004). A comparison of multiple-choice and open-ended response formats for the assessment of listening proficiency in English. *Foreign Language Annals*, 37(4), 544-53.
- Chung, J. M. (1996). The effects of using advance organizers and captions to introduce video in the foreign language classroom. *TESL Canada Journal*, 61-65.
- Chung, J. M. (1999). The effects of using video texts supported with advance organizers and captions on Chinese college students' listening comprehension: An empirical study. *Foreign Language Annals*, 32(3), 295-308.
- Chung, J. M. (2002). The effects of using two advance organizers with video texts for the teaching of listening in English. *Foreign Language Annals*, 35(2), 231-241.
- Chung, J. M., & Huang, S. C. (1998). The effects of three aural advance organizers for video viewing in a foreign language classroom. *System*, 26(4), 553-565.
- Cobb, Tom. (2008). The Compleat Lexical Tutor. Retrieved from: <https://www.lextutor.ca/> (May, 2017)
- Cohen, A. D., Oxford, R. L., & Chi, J. C. (2001). Learning style survey. Retrieved May, 2005.



- d'Ydewalle, G., & Gielen, I. (1992). Attention allocation with overlapping sound, image, and text. In Rayner K. (Eds.) *Eye movements and visual cognition* (pp. 415-427). Springer Series in Neuropsychology.
- d'Ydewalle, G., & Gielen, I. (1992). Attention allocation with overlapping sound, image, and text. In *Eye movements and visual cognition* (pp. 415-427). Springer, New York, NY.
- Danan, M. (1992). Reversed subtitling and dual coding theory: New directions for foreign language instruction. *Language learning*, 42(4), 497-527.
- Danan, M. (2004). Captioning and subtitling: undervalued language learning strategies. *Meta: Journal Des Traducteurs*, 49(1), 67-77. doi:10.7202/009021ar.
- Davey, B., & McBride, S. (1986). Effects of question-generation training on reading comprehension. *Journal of Educational Psychology*, 78(4), 256-62.
- de Vos, J. F., Schriefers, H., Nivard, M. G., & Lemhöfer, K. (2018). A meta-analysis and meta-regression of incidental second language word learning from spoken input. *Language Learning*, 68(4), 906-41.
- Del Rincón, D., Arnal, J., Latorre, A., & Sans, A. (1995). *Técnicas de investigación en ciencias sociales*. Madrid: Dykinson.
- Díaz-Cintas, J., & Remael, A. (2014). *Audiovisual Translation, Subtitling*. Routledge.
- Dörnyei, Z. (2009). *The psychology of second language acquisition*. Oxford: Oxford University Press.
- Doughty, C. (2004). Effects of instruction on learning a second language: A critique of instructed SLA research. *Form-meaning connections in second language acquisition*, 181-202.
- Douglas Fir Group. (2016). A transdisciplinary framework for SLA in a multilingual world. *The Modern Language Journal*, 100(S1), 19-47.
- d'Ydewalle, G., & De Bruycker, W. (2007). Eye movements of children and adults while reading television subtitles. *European psychologist*, 12(3), 196-205.
- d'Ydewalle, G., & Van de Poel, M. (1999). Incidental foreign-language acquisition by children watching subtitled television programs. *Journal of Psycholinguistic Research*, 28(3), 227-44. doi:10.1023/A:1023202130625
- d'Ydewalle, G., Praet, C., Verfaillie, K., & Rensbergen, J. V. (1991). Watching subtitled television: Automatic reading behavior. *Communication research*, 18(5), 650-666.

- d'Ydewalle, G., Van Rensbergen, J., & Pollet, J. (1987). Reading a message when the same message is available auditorily in another language: The case of subtitling. In *Eye movements from physiology to cognition* (pp. 313-321). Elsevier.
- Elkhafaifi, H. (2005). The effect of prelistening activities on listening comprehension in Arabic learners. *Foreign Language Annals*, 38(4), 505-13.
- Elley, W.B. (1989). Vocabulary acquisition from listening to stories. *Reading research quarterly*, 174-187.
- Ellis, N. (2013). Second language acquisition. In Trousdale, G., & T. Hoffmann (Eds.) *Oxford Handbook of Construction Grammar* (pp. 365-378). Oxford: Oxford University Press.
- Ellis, N. C. (2003). Constructions, chunking, and connectionism: The emergence of second language structure. *The handbook of second language acquisition*, 14, 63.
- Etemadi, A. (2012). Effects of bimodal subtitling of English movies on content comprehension and vocabulary recognition. *International journal of English linguistics*, 2(1), 239.
- 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. Final Report.
- European Commission (2012). *The European survey on language competences*.
- European Commission. (2017). Media use in the European Union. European Union, standard Eurobarometer 88 Project. doi:10.2775/116707
- Frumuselu, A. D. (2015). *Subtitled television series inside the efl classroom: long-term effects upon colloquial language learning and oral production* (Doctoral dissertation, Universitat Rovira i Virgili).
- Frumuselu, A.D., 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, 107-17. doi:10.1016/j.linged.2015.10.001.
- Garza, T. (1991). Evaluating the use of captioned video materials in advanced foreign language learning. *Foreign Language Annals*, 24(3), 239-58.
- Gass, S., Winke, P., Isbell, D., & Ahn, J. (2019). How captions help people learn languages: A working-memory, eye-tracking study. *Language Learning & Technology* 23(2), 84-104.
- Gesa (2019) L1/L2 subtitled TV series and EFL learning: a study on vocabulary acquisition and content comprehension at different proficiency levels.

- Gielen, I., & d'Ydewalle, G. (1989). Hoe worden ondertitelde televisieprogramma's bekeken?. *Psycholoog*.
- Gieve, S., & Clark, R. (2005). 'The Chinese approach to learning': Cultural trait or situated response? The case of a self-directed learning programme. *System*, 33(2), 261-276.
- Gilmore, A. (2007). Authentic materials and authenticity in foreign language learning. *Language Teaching*, 40(2), 97-118.
- Godfroid, A., Boers, F., & Housen, A. (2013). An eye for words: Gauging the role of attention in incidental L2 vocabulary acquisition by means of eye-tracking. *Studies in Second Language Acquisition*, 35(3), 483-517.
- Goh, C. (2000). A cognitive perspective on language learners' listening comprehension problems. *System*, 28(1), 55-75.
- Gorjian, B. (2014). The effect of movie subtitling on incidental vocabulary learning among EFL learners. *International Journal of Asian Social Science*, 4(9), 1013-1026.
- Graham, S. (2006). Listening comprehension: The learners' perspective. *System*, 34(2), 165-182.
- Gruba, P. (2004). Understanding digitized second language videotext. *Computer Assisted Language Learning*, 17(1), 51-82.
- Gruba, P. (2006). Playing the videotext: A media literacy perspective on video-mediated L2 listening. *Language Learning & Technology*, 10(2), 77-92.
- Guichon, N., & McLornan, S. (2008). The effects of multimodality on L2 learners: Implications for CALL resource design. *System*, 36(1), 85-93.
- Guillory, H. (1998). The effects of keyword captions to authentic French video on learner comprehension. *CALICO Journal*, 15(1-3), 89-108.
- Hasan, A. (2000). Learners' perceptions of listening comprehension problems. *Language, Culture and Curriculum*, 13(2), 137-53.
- 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), 181-92.
- Hirsh, D., & Nation, P. (1992). What vocabulary size is needed to read unsimplified texts for pleasure?. *Reading in a foreign language*, 8, 689-689.
- Horst, M., Cobb, T., & Meara, P. (1998). Beyond a clockwork orange: Acquiring second language vocabulary through reading. *Reading in a Foreign Language* 11(2), 207-23.

- Hsu, C.K., Hwang J.G., Chang, Y.T. & Chang, C.K. (2013). Effects of video caption modes on English listening comprehension and vocabulary acquisition using handheld devices. *Journal of Educational Technology & Society*, 16(1): 403-414. ISSN1436-4522
- Hu, M., & Nation, P. (2000). Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language*, 13, 403-30.
- Huang, H., & Eskey, D. (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.
- Hulstijn, J. H. (2003). Incidental and intentional learning. In Doughty C.J., & Long, M. (Eds.) *Handbook of Second Language Acquisition* (pp. 349-81). Malden, MA: Wiley-Blackwell.
- Hulstijn, J. H. (2013). "Incidental learning in second language acquisition". In Chapelle C.A. (Ed.) *The encyclopedia of applied linguistics* (Vol. 5, pp. 2632-40). Chichester: Wiley-Blackwell. doi:10.1002/9781405198431.wbeal0530
- Hummel, K. M. (2010). Translation and short-term L2 vocabulary retention. Hindrance or help? *Language Teaching Research* 14, 61-74. doi:10.1177/1362168809346497.
- Jones, L. (2003). Supporting listening comprehension and vocabulary acquisition with multimedia annotations: The students' voice. *CALICO Journal*, 21(1), 41-65.
- Jones, L. (2004). Testing L2 vocabulary recognition and recall using pictorial and written test items. *Language learning & technology* 8(3), 122-143. ISSN 1094-3501
- Jones, L., & Plass, J. (2002). Supporting listening comprehension and vocabulary acquisition in French with multimedia annotations. *The Modern Language Journal*, 86(4), 546-61.
- Kelley, T. (1939). The selection of upper and lower groups for the validation of test items. *Journal of Educational Psychology*, 30, 17-24.
- Khan, N., Kasdar, J., Melvin, M., Blomquist, R., Huang, E., & McEwen, J. (2015). Fresh off the Boat [TV series] Los Angeles, CA: ABC
- Kim, H. S. (2015). Using Authentic Videos to Improve EFL Students' Listening Comprehension. *International Journal of Contents*, 11(4).
- King, J. (2002). Using DVD feature films in the EFL classroom. *Computer assisted language learning*, 15(5), 509-523.
- 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.

- Koolstra, C. M., Van der Voort, T. H., & van der Kamp, L. J. T. (1997). Television's impact on children's reading comprehension and decoding skills: A 3-year panel study. *Reading Research Quarterly*, 32(2), 128-152.
- 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.
- Krashen, S. (2003). *Explorations in Language Acquisition and Use*. Portsmouth, NH: Heinemann
- Krashen, S. D. (1985). *The input hypothesis: Issues and implications*. Addison-Wesley Longman Ltd.
- Kuppens, A.H. (2010). Incidental Foreign Language Acquisition from Media Exposure. *Learning, Media and Technology* 35(1), 65-85. doi:10.1080/17439880903561876.
- Lambert, W. E., & Holobow, N. E. (1984). Combinations of printed script and spoken dialogue that show promise for students of a foreign language. *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement*, 16(1), 1.
- Latifi, M., Mobalegh, A., & Mohammadi, E. (2011). Movie subtitles and the improvement of listening comprehension ability: Does it help? *The Journal of Language Learning and Teaching*, 1(2), 18-29.
- Laufer, B. (1989). What percentage of text-lexis is essential for comprehension. *Special language: From humans thinking to thinking machines*, 316323.
- Laufer, B. (1992). How much lexis is necessary for reading comprehension?. In *Vocabulary and applied linguistics* (pp. 126-132). Palgrave Macmillan, London.
- Laufer, B. (2005). Focus on form in second language vocabulary learning. *EUROSLA Yearbook*, 5(1), 223-50.
- Laufer, B., & Ravenhorst-Kalovski, G. (2010). Lexical Threshold Revisited: Lexical Text Coverage, Learners' Vocabulary Size and Reading Comprehension. *Reading in a foreign language*, 22(1), 15-30.
- Laufer, B., & Rozovski-Roitblat, B. (2015). Retention of new words: Quantity of encounters, quality of task, and degree of knowledge. *Language Teaching Research*, 19(6), 687-711.
- Lavour, J. M., & Bairstow, D. (2011). Languages on the screen: Is film comprehension related to the viewers' fluency level and to the language in the subtitles
- Lee, J. F., & Van Patten, B. (2003). *Making communicative language teaching happen*. New York, NY: McGraw Hill.

- Lee, S. (2007). Effects of textual enhancement and topic familiarity on Korean EFL students' reading comprehension and learning of passive form. *Language Learning*, 57(1), 87-118
- Lekkai, I. (2014). Incidental foreign-language acquisition by children watching subtitled television programs. *TOJET: The Turkish Online Journal of Educational Technology* 13(4), 81-87
- Lertola, J. (2012). The effect of the subtitling task on vocabulary learning. *Translation research projects*, 4, 61-70.
- Li, C. (2014). An alternative to language learner dependence on L2 caption-reading input for comprehension of sitcoms in a multimedia learning environment. *Journal of Computer Assisted Learning*, 30(1), 17-29.
- Lindgren, E., & Muñoz, C. (2013). The Influence of Exposure, Parents, and Linguistic Distance on Young European Learners Foreign Language Comprehension. *International Journal of Multilingualism* 10(1), 105-29. doi:10.1080/14790718.2012.679275.
- Lwo, L., & Lin, M. (2012). The Effects of Captions in Teenagers' Multimedia L2 Learning. *ReCALL*, 24(2), 188-208. doi:10.1017/s0958344012000067.
- Maleki, A., & Safaee Rad, M. (2011). The effect of visual and textual accompaniments to verbal stimuli on the listening comprehension test performance of Iranian high and low proficient EFL learners. *Theory and Practice in Language Studies*, 1(1), 28-36.
- Malone, J. (2018). Incidental Vocabulary Learning in SLA. *Studies in Second Language Acquisition*, 1-25. doi:10.1017/s0272263117000341.
- Mardani, M., & Najmabadi, A. (2016). The Effect of Movies with Different Types of Subtitles on Incidental English Vocabulary Learning of Iranian High School Students. *Research in English Language Pedagogy*, 4(1), 52-63.
- Markham, P. (1999). Captioned videotapes and second-language listening word recognition. *Foreign Language Annals*, 32(3), 321-28.
- Markham, P., & Peter, L. (2003). The influence of English language and Spanish language captions on foreign language listening/reading comprehension. *Journal of Educational Technology Systems*, 31(3), 331-41.
- Markham, P., Peter, L., & McCarthy, T. (2001). The effects of native language vs. target language captions on foreign language students' DVD video comprehension. *Foreign Language Annals*, 34(5), 439-45.
- Matielo, R., Collet, T., & D'Ely, R.C.S.F. (2013). The effects of interlingual and intralingual subtitles on vocabulary learning by Brazilian EFL learners: an

- exploratory study. *Intercâmbio. Revista do Programa de Estudos Pós-Graduados em Linguística Aplicada e Estudos da Linguagem* 27, 83-99. ISSN2237-759x.
- Matielo, R., de Oliveira, R., & Baretta, L. (2017). Intralingual subtitles, interlingual subtitles, and video comprehension: insights from an exploratory study. *Letrônica*, 10(2), 758-74.
- Mayer, R. (2001). *Multimedia learning*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9781139164603>
- Mayer, R. (2014). *The Cambridge handbook of multimedia learning*. Cambridge: Cambridge University Press.
- Mayer, R., & Moreno, R. (1998). A split-attention effect in multimedia learning: Evidence for dual processing systems in working memory. *Journal of Educational Psychology*, 90(2), 312–320. <https://doi.org/10.1037/0022-0663.90.2.312>
- Mayer, R., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist*, 38(1), 43-52.
- Meara, P., & Jones, G. (1988). Vocabulary Size as a Placement Indicator.
- Meara, P., & Milton, J. (2003). *X\_Lex: the Swansea levels test*. Berkshire: Express Publishing.
- Meskill, C. (1996). Listening skills development through multimedia. *Journal of Educational Multimedia and Hypermedia*, 5(2), 179-201.
- Milton, J. (2010). The development of vocabulary breadth across the CEFR levels. *Communicative proficiency and linguistic development: Intersections between SLA and language testing research*, 211-232.
- Miralpeix, I. (2009). Modelling and Assessing Vocabulary Development, Helmut Daller, James Milton, Jeanine Treffers-Daller (Eds.), Cambridge University Press, Cambridge (2007). xiii+ 274 pp.
- Miralpeix, I. (2012, March). *X\_Lex and Y\_Lex: A validation study*. Paper presented at the 22nd Lexical Studies Conference, Newtown, UK.
- Mitterer, H., & McQueen, J. (2009). Foreign subtitles help but native-language subtitles harm foreign speech perception. *PloS one*, 4(11): e7785.
- Mohd Jelani, N.A., and F. Boers. 2018. "Examining Incidental Vocabulary Acquisition from Captioned Video." *Approaches to Learning, Testing, and Researching L2 Vocabulary ITL - International Journal of Applied Linguistics* 169 (1): 169-90. doi:10.1075/itl.00011.jel.

- Montero-Perez, M., Peters, E., & Desmet, P. (2014). Is less more? Effectiveness and perceived usefulness of keyword and full captioned video for L2 listening comprehension. *ReCALL*, 26(1), 21-43.
- 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.
- 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), 118-41. doi:10125/44357
- Montero-Perez, M., Van Den Noortgate, W., & Desmet, P. (2013). Captioned video for L2 listening and vocabulary learning: A meta-analysis. *System*, 41(3), 720-39. doi:10.1016/j.system.2013.07.013.
- Muñoz, C. (2017). The role of age and proficiency in subtitle reading. An eye-tracking study. *System* 67, 77-86. DOI: 10.1016/j.system.2017.04.015
- Muñoz, C. (forthcoming) Boys like games and girls like movies. Age and gender differences in out-of-school contact with English.
- Muñoz, C., Cadierno, T. & Casas, I. (2018). Different starting points for English language learning: a comparative study of Danish and Spanish young learners. *Language Learning* 68. DOI: 10.1111/lang.12309
- Naghizadeh, M., & Darabi, T. (2015). The Impact of Bimodal, Persian and No-Subtitle Movies on Iranian EFL Learners' L2 Vocabulary Learning. *Journal of Applied Linguistics and Language Research* 2(2), 66-79. ISSN2376-760x
- Nagira, Y. (2011). Vocabulary learning through captions. In *International Conference on Languages, Literature and Linguistics* (Vol. 26, pp. 95-99).
- Nation, I. S. (2001). *Learning vocabulary in another language*. Ernst Klett Sprachen.
- Nation, P. (2006). How large a vocabulary is needed for reading and listening? *Canadian modern language review*, 63(1), 59-82. doi:10.3138/cmlr.63.1.59
- Nation, P. (2007). The four strands. *International Journal of Innovation in Language Learning and Teaching*, 1(1), 2-13. doi: 10.2167/illt039.0
- Nation, P. (2012). The BNC/COCA Word Family Lists. *Document bundled with Range program with BNC/COCA lists*, 25.
- Nation, P. (2015). Principles guiding vocabulary learning through extensive reading. *Reading in a Foreign Language* 27 (1): 136-145. URL: <http://nflrc.hawaii.edu/rfl>



- Nation, P., & Heatley, A. (2002). Range: A program for the analysis of vocabulary in texts [software]. Retrieved from: <https://www.victoria.ac.nz/lals/about/staff/paul-nation>
- Nation, P., & Newton, J. (2009). *Teaching ESL/EFL listening and speaking. ESL & Applied linguistics professional series*. Abingdon: Routledge.
- Nation, P., & Webb, S. A. (2011). *Researching and analyzing vocabulary*. Boston, MA: Heinle, Cengage Learning.
- Neuman, S., & 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.
- Noblitt, J. S. (1995). Cognitive approaches to listening comprehension. *Dimension*, 95.
- Noreillie, S., Kestemont, B., Heylen, K., Desmet, P., & Peters, E. (2018). Vocabulary knowledge and listening comprehension at an intermediate level in English and French as foreign languages. *ITL-International Journal of Applied Linguistics*, 169(1), 212-31.
- Nurmukhamedov, U. (2017). Lexical coverage of TED Talks: Implications for vocabulary instruction. *TESOL Journal*, 8(4), 768-90.
- Nurmukhamedov, U., & Webb, S. (2019). Lexical coverage and profiling. *Language Teaching*, 52(2), 188-200.
- Ockey, G. (2007). Construct implications of including still image or video in computer-based listening tests. *Language Testing*, 24(4), 517-37.
- Ortega, L. (2014). *Understanding second language acquisition*. Routledge.
- Paivio, A. (1986). *Mental representations: a dual coding approach*. Oxford: Oxford University Press.
- Paivio, A. (2008). Mind and its evolution: a dual coding theoretical approach. *Psychol Rec*, 59, 295-300.
- Park, G. P. (2004). Comparison of L2 listening and reading comprehension by university students learning English in Korea. *Foreign Language Annals*, 37(3), 448-58.
- Pavakanun, U., & d'Ydewalle, G. (1992). Watching foreign television programs and language learning. In Engel, F. L., Bouwhuis, D. G., Bosser, T., & d'Ydewalle, G. (Eds.), *Cognitive modelling and interactive environments in language learning* (pp. 193-98). Berlin: Springer.

- Pearson, P., & Johnson, D. (1978) *Teaching reading comprehension*. New York: Holt, Rinehart, & Winston.
- Pellicer-Sánchez, A. (2016). Incidental L2 vocabulary acquisition from and while reading: An eye-tracking study. *Studies in Second Language Acquisition*, 38(1), 97-130.
- 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.
- Perego, E., Del Missier, F., Porta, M., & Mosconi, M. (2010). The cognitive effectiveness of subtitle processing. *Media psychology*, 13(3), 243-272.
- Peters, E. (2017). The effect of out-of-class exposure on learners' vocabulary size. In AAAL, Date: 2017/03/18-2017/03/21, Location: Portland, Oregon (USA).
- Peters, E. (2018). The Effect of Out-of-class Exposure to English Language Media on Learners' Vocabulary Knowledge. *Approaches to Learning, Testing, and Researching L2 Vocabulary ITL - International Journal of Applied Linguistics* 169(1), 142-68. doi:10.1075/itl.00010.pet.
- Peters, E. (2019). The effect of imagery and subtitles on L2 vocabulary learning from audio-visual input. In AAAL, Date: 2019/03/09-2019/03/12, Location: Atlanta.
- Peters, E., & Webb, S. (2018). Incidental vocabulary acquisition through viewing L2 television and factors that affect learning. *Studies in Second Language Acquisition*, 40(3), 551-77. doi:10.1017/s0272263117000407.
- Peters, E., Heynen, E., & Puimège, E. (2016). Learning Vocabulary through Audiovisual Input: The Differential Effect of L1 Subtitles and Captions. *System* 63, 134-48. doi:10.1016/j.system.2016.10.002.
- Peters, E., Noreillie, A. S., Heylen, K., Bulté, B., & Desmet, P. (2019). The Impact of Instruction and Out-of-School Exposure to Foreign Language Input on Learners' Vocabulary Knowledge in Two Languages. *Language Learning*.
- Pigada, M., & Schmitt, N. (2006). Vocabulary acquisition from extensive reading: A case study. *Reading in a foreign language*, 18(1), 1-28.
- Plass, J., & Jones, L. (2005). Multimedia learning in second language acquisition. *The Cambridge handbook of multimedia learning*, 467-488.
- Price, K. (1983). Closed-captioned TV: An untapped resource. *Matsol Newsletter*, 12(2), 1-8.

- Puimège, E., & Peters, E. (2019). Learning L2 vocabulary from audiovisual input: an exploratory study into incidental learning of single words and formulaic sequences. *The Language Learning Journal*, 47(4), 424-438.
- Pujadas, G. & Muñoz, C. (2017, April) *Learning through subtitles. Learners' preferences and task perception*. Paper presented at the Task-Based Language Teaching Conference, Barcelona, Spain
- Pujadas, G. & Muñoz, C. (2018, October) *What words are learnt better through TV series? The effects of age, proficiency and type of instruction*. Paper presented at the Second Language Research Forum, Mont-real, Canada
- Pujadas, G. & Muñoz, C. (in prep) Does pre-teaching enhance vocabulary learning while watching TV? Evidence from an eye-tracking study with adolescent EFL learners.
- Pujolà, J. T. (2002). CALLing for help: Researching language learning strategies using help facilities in a web-based multimedia program. *ReCALL*, 14(2), 235-262.
- Rice, M.L., Huston, A.C., Truglio, R. & Wright, J.C. (1990). Words from "Sesame Street: Learning Vocabulary While Viewing. *Developmental Psychology* 26(3), 421-28. doi:10.1037//0012-1649.26.3.421.
- Robinson, P. (2003). Attention and memory during SLA. *The handbook of second language acquisition*, 631-678.
- Robinson, P., Mackey, A., Gass, S. M., & Schmidt, R. (2012). Attention and awareness in second language acquisition. *The Routledge handbook of second language acquisition*, 247-267.
- Rodgers, M. (2013). *English language learning through viewing television: An investigation of comprehension, incidental vocabulary acquisition, lexical coverage, attitudes, and captions* (Unpublished doctoral dissertation). Victoria University of Wellington.
- Rodgers, M. (2016). Extensive listening and viewing: the benefits of audiobooks and television. *The European Journal of Applied Linguistics and TEFL*, 5(2), 43-57.
- Rodgers, M. (2018a). The images in television programs and the potential for learning unknown words. *Approaches to Learning, Testing, and Researching L2 Vocabulary ITL - International Journal of Applied Linguistics* 169(1), 191-211. doi:10.1075/itl.00012.rod.
- Rodgers, M. (2018b, October). *The effects of captions on facets of viewing comprehension of an authentic television program*. Paper presented at the Second Language Research Forum, Montreal, Canada.

- Rodgers, M. (2019, July) *Game on for comprehensible input and vocabulary learning: The lexical demands of videos games*. Paper presented at the Vocab@Leuven Conference, Leuven, Belgium.
- Rodgers, M., & Webb, S. (2011). Narrow viewing: The vocabulary in related television programs. *TESOL Quarterly*, 45(4), 689-717. doi:10.5054/tq.2011.268062
- Rodgers, M., & Webb, S. (2017). The effects of captions on EFL learners' comprehension of English-language television programs. *CALICO Journal*, 34(1), 20-38.
- 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.
- Schmitt, N. (2008). Instructed second language vocabulary learning. *Language teaching research*, 12(3), 329-363.
- Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. Springer.
- Schmitt, N., & Carter, R. (2000). The lexical advantages of narrow reading for second language learners. *TESOL Journal*, 9, 4-9.
- Shohamy, E. (1984). Does the testing method make a difference? The case of reading comprehension. *Language testing*, 1(2), 147-70.
- Shohamy, E., & Inbar, O. (1991). Validation of listening comprehension tests: The effect of text and question type. *Language testing*, 8(1), 23-40. doi:10.1111/j.1944-9720.2004.tb02701.x.
- Sockett, G. (2014). *The online informal learning of English*. Springer.
- Stæhr, L. S. (2009). Vocabulary knowledge and advanced listening comprehension in English as a foreign language. *Studies in second language acquisition*, 31(4), 577-607.
- Statistics Canada (2005). *Average time spent on activities, by sex*. Ottawa, ON: Statistics Canada. Retrieved from: <http://www40.statcan.ca/l01/cst01/famil36a.htm> (June 2019).
- Stewart, M., & Pertusa, I. (2004). Gains to language learners from viewing target language closed-captioned films. *Foreign language annals*, 37(3), 438-42.
- Sweller, J. (1999). *Instructional design in technical areas*. Camberwell, Australia: ACER Press.
- Sweller, J. (2005). Implications of cognitive load theory for multimedia learning. *The Cambridge handbook of multimedia learning*, 19-30.

- Sydorenko, T. (2010). Modality of input and vocabulary acquisition. *Language Learning and Technology*, 14(2), 50-73. doi:10125/44214
- Taylor, G. (2005). Perceived processing strategies of students watching captioned video. *Foreign Language Annals*, 38(3), 422-27.
- 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), 559-99.
- United States Department of Labor (2006). *American time use survey summary*. Washington, DC: United States Department of Labor. Retrieved from: [http://www.bls.gov/news.release/archives/atus\\_06032008.htm](http://www.bls.gov/news.release/archives/atus_06032008.htm) (June 2019).
- Van Heuven, W. J., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *The Quarterly Journal of Experimental Psychology*, 67(6), 1176-1190.
- Van Lommel, S., Laenen, A., & d'Ydewalle, G. (2006). Foreign-grammar acquisition while watching subtitled television programmes. *British Journal of Educational Psychology*, 76(2), 243-258.
- van Zeeland, H., & Schmitt, N. (2012). Lexical coverage in L1 and L2 listening comprehension: The same or different from reading comprehension? *Applied Linguistics*, 34(4), 457-79.
- van Zeeland, H., & Schmitt, N. (2013). Incidental vocabulary acquisition through L2 listening: A dimensions approach. *System*, 41(3), 609-624.
- Vandergrift, L. (2007). Recent developments in second and foreign language listening comprehension research. *Language teaching*, 40(3), 191-210.
- Vandergrift, L. (2011). Second language listening: Presage, process, product, and pedagogy. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 455–471). New York, NY: Routledge.
- Vanderplank, R. (1988). The value of teletext sub-titles in language learning. *ELT Journal*, 42(4), 272–281.
- 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. (2014). Listening and understanding. In *Debates in modern languages education* (pp. 73-85). Routledge.

- Vanderplank, R. (2016a). 'Effects of' and 'effects with' captions: How exactly does watching a TV programme with same-language subtitles make a difference to language learners? *Language Teaching*, 49(2), 235-50.  
doi:10.1017/s0261444813000207.
- Vanderplank, R. (2016b). *Captioned media in foreign language learning and teaching: Subtitles for the deaf and hard-of-hearing as tools for language learning*. New York: Springer.
- Vanderplank, R. (2019). 'Gist watching can only take you so far': attitudes, strategies and changes in behaviour in watching films with captions. *The Language Learning Journal*, 1-17.
- VanPatten, B. (1996). *Input processing and grammar instruction: Theory and research*. New Jersey: Ablex.
- VanPatten, B. (2002). Processing instruction: An update. *Language Learning*, 52(4), 755-803.
- Vidal, K. (2003). Academic listening: A source of vocabulary acquisition? *Applied Linguistics*, 24(1), 56-89.
- Vidal, K. (2011). A comparison of the effects of reading and listening on incidental vocabulary acquisition. *Language Learning*, 61(1), 219-258.
- Vulchanova, M., Aurstad, L. M., Kvitnes, I. E., & Eshuis, H. (2015). As naturalistic as it gets: subtitles in the English classroom in Norway. *Frontiers in Psychology*, 5, 1510.
- Wagner, E. (2002). Video listening tests: A pilot study. *Teachers College, Columbia University Working Papers in TESOL & Applied Linguistics*, 2(1), 1-39.
- 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), 130-63.
- Webb, S. (2007). The effects of repetition on vocabulary knowledge. *Applied linguistics*, 28(1), 46-65.
- Webb, S. (2010a). A corpus driven study of the potential for vocabulary learning through watching movies. *International Journal of Corpus Linguistics*, 15(4), 497-519.
- Webb, S. (2010b). Pre-learning low-frequency vocabulary in second language television programmes. *Language Teaching Research*, 14(4), 501-15.  
doi:10.1177/1362168810375371
- Webb, S. (2010c). Using Glossaries to Increase the Lexical Coverage of Television Programs. *Reading in a Foreign Language*, 22(1), 201-221.

- Webb, S. (2011). Selecting television programs for language learning: Investigating television programs from the same genre. *International Journal of English Studies*, 11(1), 117-35.
- Webb, S. (2012). Repetition in incidental vocabulary learning. *The encyclopedia of applied linguistics*, 1-6.
- Webb, S. (2015). Extensive viewing: Language learning through watching television. In Nunan, D., & Richards, J. (Eds.) *Language learning beyond the classroom* (pp. 175-184). Abingdon: Routledge.
- Webb, S., & Chang, A. C. (2015a). How does prior word knowledge affect vocabulary learning progress in an extensive reading program? *Studies in Second Language Acquisition*, 37, 651–675. <https://doi.org/10.1017/S0272263114000606>
- Webb, S., & Chang, A. C. (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.
- Webb, S., & Nation, P. (2017). *How vocabulary is learned*. Oxford: Oxford University Press.
- Webb, S., & Rodgers, M. (2009a). Vocabulary demands of television programs. *Language Learning*, 59(2), 335-66. doi:10.1111/j.1467-9922.2009.00509.x.
- Webb, S., & Rodgers, M. (2009b). The lexical coverage of movies. *Applied Linguistics*, 30(3), 407-27. doi:10.1093/applin/amp010.
- Webb, S., Newton, J., & Chang, A. (2013). Incidental learning of collocation. *Language Learning*, 63(1), 91-120.
- Winke, P., Gass, S., & Sydorenko, T. (2010). The effects of captioning videos used for foreign language listening activities. *Language Learning and Technology*, 14(1), 65-86. doi:10125/44203
- Yang, H. Y. (2014). The effects of advance organizers and subtitles on EFL learners' listening comprehension skills. *Calico Journal*, 31(3), 345-373.
- Yoshino, S., Kano, N., & Akahori, K. (2000). The effects of English and Japanese captions on the listening comprehension of Japanese EFL students. *Language Laboratory*, 37, 111-130.
- Zarei, A. (2008). The effect of bimodal, standard, and reversed subtitling on L2 vocabulary recognition and recall. *Pazhuhesh-e Zabanha-ye Khareji* 49, 65-84.

Zarei, 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

Zarei, A., & Saddeghi, M. (2011). The effects of synchronous and asynchronous interlingual and intralingual transcript presentation on L2 vocabulary comprehension and production. *Journal of Teaching English Language and Literature Society of Iran*, 5, 101-123.





## Appendices

---

Appendix A. 1. Lexical coverage per episode

Episode	Tokens														TOTAL
	1,000 frequency band		2,000 frequency band		3,000 frequency band		Proper nouns		Marginal Words		TIs beyond 3,000 band		4,000 + frequency band		
	n	%	n	%	n	%	n	%	n	%	n	%	n	%	
1	2851	87.4	104	3.19	48	1.47	137	4.20	14	0.43	2	0.06	106	3.25	3262
2	2399	87.4	106	3.86	41	1.49	54	1.97	43	1.57	17	0.62	85	3.10	2745
3	2208	83.73	113	4.29	62	2.35	101	3.83	41	1.55	7	0.27	105	3.98	2637
4	2527	85.37	130	4.39	52	1.76	91	3.07	57	1.93	8	0.27	95	3.21	2960
5	2516	85.14	132	4.47	44	1.49	84	2.84	40	1.35	17	0.58	122	4.13	2955
6	2602	85.45	132	4.33	66	2.17	63	2.07	64	2.10	27	0.89	91	2.99	3045
7	2712	84.94	122	3.82	50	1.57	135	4.23	54	1.69	6	0.19	114	3.57	3193
8	2770	87.52	96	3.03	56	1.77	72	2.28	47	1.49	17	0.54	107	3.38	3165
9	2436	83.28	128	4.38	46	1.57	113	3.86	61	2.09	20	0.68	121	4.14	2925
10	2605	85.13	116	3.79	69	2.25	90	2.94	45	1.47	7	0.23	128	4.18	3060
11	2395	84.72	123	4.35	40	1.41	67	2.37	64	2.26	6	0.21	132	4.67	2827
12	2307	84.26	148	5.41	39	1.42	104	3.80	41	1.50	1	0.04	98	3.58	2738
13	2322	84.53	139	5.06	37	1.35	74	2.69	38	1.38	9	0.33	128	4.66	2747
14	2178	85.31	93	3.64	48	1.88	64	2.51	43	1.69	20	0.78	107	4.19	2553
15	2607	86.96	116	3.87	41	1.37	94	3.14	29	0.97	11	0.37	100	3.34	2998
16	2283	83.2	121	4.41	67	2.44	86	3.13	46	1.68	12	0.44	129	4.70	2744
17	2493	87.38	99	3.47	40	1.4	73	2.56	45	1.58	15	0.53	88	3.09	2853
18	2195	81.36	102	3.78	83	3.08	126	4.67	46	1.71	4	0.15	142	5.26	2698
19	2771	86.16	105	3.26	57	1.77	85	2.64	38	1.18	17	0.53	143	4.45	3216
20	2116	82.05	149	5.78	39	1.51	132	5.12	46	1.78	2	0.08	95	3.68	2579
21	2336	83.46	149	5.32	55	1.96	74	2.64	58	2.07	17	0.61	110	3.93	2799
22	2672	84.8	116	3.68	67	2.13	102	3.24	91	2.89	1	0.03	102	3.24	3151
23	2274	82.48	104	3.77	68	2.47	117	4.24	28	1.02	17	0.62	149	5.40	2757
24	2342	85.38	122	4.45	73	2.66	67	2.44	28	1.02	11	0.40	100	3.65	2743
All	58917	84.89	2865	4.16	1288	1.86	2205	3.19	1107	1.60	271	0.39	2697	3.91	69350

Appendix A. 2. List of target items

Target item	Spanish equivalent (as in subtitles)	Term	Session	Session in term	Recency	Spacing	Frequency in target episode	Total frequency in term	PoS	BNC/COCA word family lists	SubtLex	Concret- ness
hell	demonios!	1	1	1	7	2	4	13	4	1	470.82	-
janitor	conserje	1	1	1	3	2	2	3	1	16	5.73	4.68
to figure out	descifrar	1	1	1	3	2	3	6	2	-	-	1.93
to hire	contratar	1	1	1	7	2	2	4	2	2	25.53	2.59
to move	mudarse	1	1	1	8	2	8	20	2	1	418.14	3.25
cheers	salud!	1	2	2	2	1	3	3	4	2	22.71	-
crouton	tostón	1	2	2	2	1	8	8	1	18	0.25	4.9
jukebox	gramola	1	2	2	2	1	3	3	1	10	2.27	4.93
napkin	servilleta	1	2	2	2	1	6	6	1	7	3.61	4.93
tough	duro	1	2	2	8	2	3	11	3	2	90.51	2.5
nightmare	pesadilla	1	3	3	4	2	2	3	1	3	22.39	2.96
rib	costilla	1	3	3	3	1	6	6	1	4	5.9	4.9
to fit in	encajar	1	3	3	3	2	3	3	2	-	-	2.17
to hand out	repartir	1	3	3	3	2	2	3	2	-	-	3.46
to struggle	esforzarse	1	3	3	4	1	2	6	2	2	13.37	2.79
curly	rizado	1	4	4	4	1	2	2	3	3	5.76	4.41
cushy	acomodado/a	1	4	4	4	1	3	3	3	15	0.71	2.34
fake	artificial	1	4	4	4	1	3	3	3	5	36.33	1.97
gross	asqueroso	1	4	4	4	2	2	3	3	3	15.27	2.96
to gloat	deleitarse	1	4	4	4	1	2	2	2	9	1.57	2.3
AC	aire acondicionado	1	5	5	5	1	10	10	1	-	2.16	4.21
mall	centro comercial	1	5	5	8	2	5	10	1	7	18.9	4.83
real estate	inmobiliaria	1	5	5	5	1	4	4	1	-	0.02	4.25

Target item	Spanish equivalent (as in subtitles)	Term	Session	Session in term	Recency	Spacing	Frequency in target episode	Total frequency in term	PoS	BNC/COCA word family lists	SubtLex	Concret- ness
to handle	manejar	1	5	5	7	2	3	7	2	2	108.41	4.57
to loiter	pasar tiempo	1	5	5	5	1	2	2	2	9	0.25	3.61
billboard	cartel	1	6	6	6	1	10	10	1	9	1.35	4.83
buckle	hebilla	1	6	6	6	1	2	2	1	9	5.04	4.92
carpool	coche compartido	1	6	6	6	1	6	6	1	-	0.71	3.9
franchise	franquicia	1	6	6	6	1	6	6	1	5	2.37	3.72
knockoff	copia	1	6	6	6	1	3	3	1	-	0.45	2.85
hedghehog	erizo	1	7	7	7	1	2	2	1	6	0.29	4.93
Jewish	judío	1	7	7	7	1	4	4	3	4	13.76	3.32
principal	director	1	7	7	7	1	2	2	1	3	13.75	4.79
ride	llevar en coche	1	7	7	7	2	4	8	1	2	135.37	3.75
to fire	despedir	1	7	7	7	1	11	11	2	1	215.49	2.48
hairdryer	secador de pelo	1	8	8	8	1	2	2	1	-	0.22	4.97
realtor	agente inmobiliario	1	8	8	8	2	5	7	1	19	1.8	4.61
shield	escudo	1	8	8	8	1	4	4	1	3	8.2	4.66
to ace	aprobar	1	8	8	8	1	2	2	2	5	14.73	-
to quit	renunciar	1	8	8	8	1	5	5	2	5	90.1	2.56
chickenpox	varicela	2	9	1	1	1	8	8	1	-	0.45	4.56
guest	invitado	2	9	1	1	1	3	3	1	3	39.94	3.83
jealous	celoso	2	9	1	1	2	12	12	3	4	38.27	2.17
straight	heterosexual	2	9	1	1	1	2	2	3	1	-	-
to date	salir (con alguien)	2	9	1	1	1	9	11	2	1	11.73	2.32
necklace	collar	2	10	2	2	1	3	3	1	6	9.75	4.96
rug	alfombra	2	10	2	2	1	4	4	1	4	10.41	4.79
to cash	cobrar	2	10	2	4	2	4	5	1	2	72.43	-
to ground	castigar	2	10	2	2	1	4	4	2	1	7.78	2.07

Target item	Spanish equivalent (as in subtitles)	Term	Session	Session in term	Recency	Spacing	Frequency in target episode	Total frequency in term	PoS	BNC/COCA word family lists	SubtLex	Concret- ness
to trip	tropezar	2	10	2	2	1	5	5	2	2	-	-
coach	entrenador	2	11	3	3	2	6	7	1	2	47.63	4.12
drill	ejercicio	2	11	3	3	1	3	3	1	3	13.75	4.4
play	obra de teatro	2	11	3	3	1	12	12	1	1	-	-
teamwork	trabajo en equipo	2	11	3	3	1	3	3	1	10	2	2.37
to bounce	botar	2	11	3	3	1	4	4	2	3	9.84	3.86
cool	guay	2	12	4	8	2	5	19	3	2	195.88	3.53
fridge	nevera	2	12	4	4	1	4	4	1	2	9.84	4.92
proud	orgulloso/a	2	12	4	5	2	2	5	3	2	83.63	2.07
to join	apuntarse (a algo)	2	12	4	4	1	5	5	2	1	83.43	2.86
to trade	intercambiar	2	12	4	4	1	3	3	2	1	-	-
bill	factura	2	13	5	5	1	5	5	1	1	118.45	4.68
fee	tarifa	2	13	5	5	2	9	11	1	2	9.69	3.89
grade	curso	2	13	5	6	2	5	9	1	2	29.27	3
ride	atracción	2	13	5	5	1	5	5	1	2	135.37	3.75
whale	ballena	2	13	5	5	1	3	3	1	4	11.25	4.96
cabbage	collar	2	14	6	6	1	2	2	1	5	2.9	4.75
chess	ajedrez	2	14	6	6	1	6	6	1	6	7.45	4.7
kilt	falda escocesa	2	14	6	6	1	2	2	1	7	0.61	4.71
piccolo	flautín	2	14	6	6	1	9	9	1	16	1.27	4.71
to hang out	pasar el rato	2	14	6	6	1	3	3	2	-	-	2.59
deal	acuerdo	2	15	7	7	2	8	11	1	1	261.37	3.1
dump	vertedero	2	15	7	7	1	3	3	1	2	28.82	3.93
floor mat	alfombrilla	2	15	7	7	1	7	7	1	4	-	-
stuffed animal	peluche	2	15	7	7	1	2	2	1	-	-	-
van	furgoneta	2	15	7	7	2	3	4	1	2	51.78	4.72

Target item	Spanish equivalent (as in subtitles)	Term	Session	Session in term	Recency	Spacing	Frequency in target episode	Total frequency in term	PoS	BNC/COCA word family lists	SubtLex	Concret- ness
bow	lazo	2	16	8	8	2	3	5	1	3	20.27	4.61
costume	disfraz	2	16	8	8	2	8	10	1	4	14.14	4.57
neighbourhood	barrio	2	16	8	8	2	4	6	1	2	36.69	4.75
prank	broma	2	16	8	8	1	2	2	1	11	4	2.9
scary	escalofriante	2	16	8	8	2	5	7	3	2	26.2	3
noodle	fideos	3	17	1	1	1	8	8	1	8	6.06	4.96
sleepover	fiesta de pijamas	3	17	1	1	1	4	4	1	-	1.45	3.39
to trust	confiar	3	17	1	1	1	2	2	2	1	178.18	2.04
trap	trampa	3	17	1	1	1	2	2	1	3	23.84	4.3
weird	raro	3	17	1	7	2	5	9	3	2	101.1	1.59
cranberry	arándano	3	18	2	4	2	3	4	1	11	1.94	4.96
hen	gallina	3	18	2	7	2	6	9	1	3	3.2	4.9
oven	horno	3	18	2	2	1	4	4	1	3	8.88	4.97
raw	crudo/a	3	18	2	2	1	5	5	3	3	10.18	3.35
turkey	pavo	3	18	2	2	1	7	7	1	3	22.61	4.89
fortune teller	vidente	3	19	3	3	1	5	5	1	-	0.25	4.15
friendship	amistad	3	19	3	3	1	5	5	1	1	22.82	2.39
orchids	orquídea	3	19	3	3	1	2	2	1	9	2.63	4.92
to cut off	desconectar	3	19	3	4	2	4	5	2	-	-	3.39
to share	compartir	3	19	3	8	2	3	8	2	1	69.51	2.96
bakery	panadería	3	20	4	4	1	6	6	1	2	3.51	4.83
binder	archivador	3	20	4	4	1	3	3	1	2	0.76	4.89
Christmas Eve	Nochebuena	3	20	4	4	1	3	3	1	3	-	-
enough	suficiente	3	20	4	8	2	5	15	4	1	501.33	1.33
to make sense	tener sentido	3	20	4	5	2	5	7	2	-	-	2.11
envelope	sobre	3	21	5	5	1	14	14	1	3	10.06	4.93

Target item	Spanish equivalent (as in subtitles)	Term	Session	Session in term	Recency	Spacing	Frequency in target episode	Total frequency in term	PoS	BNC/COCA word family lists	SubtLex	Concreteness
fireworks	fuegos artificiales	3	21	5	5	1	7	7	1	5	5.63	4.86
haircut	corte de pelo	3	21	5	5	1	2	2	1	6	8.41	4.77
to mess up	equivocarse	3	21	5	5	1	3	3	2	-	-	2.93
wealth	riqueza	3	21	5	5	1	5	5	1	4	6.35	2.63
tax	impuesto	3	22	6	8	2	4	8	1	1	14.55	3.89
to bet	apostar	3	22	6	6	2	3	6	2	1	171.82	3.1
to dare	retar	3	22	6	6	1	6	6	2	2	55.41	2.28
to rush	apresurarse	3	22	6	6	1	2	2	2	2	5.25	2
wall	pared	3	22	6	6	2	4	6	1	1	70.69	4.86
dull	aburrido	3	23	7	7	1	6	6	3	3	12.08	2.37
oyster	ostra	3	23	7	7	1	3	3	1	7	3.06	4.85
review	crítica	3	23	7	7	1	7	7	1	4	14.8	2.81
threat	amenaza	3	23	7	7	1	2	2	1	2	20.76	2.63
to tease	burlarse	3	23	7	7	1	2	2	2	3	5.69	2.59
bunk	litera	3	24	8	8	1	5	5	1	5	6.27	4.3
flight attendant	asistente de vuelo	3	24	8	8	1	8	8	1	-	-	4.89
nerd	friki	3	24	8	8	1	4	4	1	16	3.29	3.23
to gamble	apostar	3	24	8	8	1	2	2	2	3	8.94	3.17
to pick on	meterse (con alguien)	3	24	8	8	1	3	3	2	-	-	-

**Term:** 1 = Term 1; 2 = Term 2; 3 = Term 3; **Session:** number of session within the overall intervention (out of 24); **Session in term:** number of session within the term, from pre- to post-test (from 1 to 8)

**Recency:** last session in which the TIs was encountered within the term (from 1 to 8); **Spacing:** 1 = massed (all encounters within the same episode); 2 = spaced (TI encountered in more than one episode)

**Frequency in episode:** number of encounters within the target episode (i.e. when the TI was pre-taught in the focused groups); **Frequency in term:** total number of encounters with the TI from pre- to post-test

**PoS:** part of speech; 1 = nouns; 2 = verbs (including phrasal verbs); 3 = adjectives; 4 = others (expressions, multiword units)

**BNC/COCA word family list:** word family in the BNC/COCA word list corpus (e.g. 1 = 1,000 family band); **SubtLex:** frequency per million in the SubtLex-US corpus

**Concreteness:** concreteness ratings by Brysbaert, Warriner and Kuperman (2014)



Appendix A. 3. List of non-target items in the vocabulary post-viewing task

Distractors	Translation	Term	Session	Session in term	Freq. in target episode	Total freq. in term	PoS	BNC / COCA word-list
skillet	<i>sartén</i>	1	5	5	2	2	1	17
to dock	<i>descontar</i>	1	5	5	2	2	2	2
to earn	<i>ganar</i>	1	5	5	2	2	2	2
smart	<i>inteligente</i>	1	6	6	3	5	3	2
to babysit	<i>hacer de canguro</i>	1	6	6	3	3	2	6
to steal	<i>robar</i>	1	6	6	2	6	2	3
host	<i>anfitrión</i>	1	7	7	2	3	1	3
awesome	<i>genial</i>	1	7	7	2	5	2	10
cop	<i>policía</i>	1	7	7	3	4	1	3
detention	<i>castigo</i>	1	8	8	2	2	1	5
houseboat	<i>casa flotante</i>	1	8	8	2	2	1	-
congratulations	<i>felicidades</i>	1	8	8	4	4	4	3
blind spot	<i>punto ciego</i>	2	9	1	3	3	4	-
clue	<i>pista</i>	2	9	1	2	2	1	2
science fair	<i>feria de ciencias</i>	2	9	1	10	10	1	-
bull	<i>toro</i>	2	10	2	7	7	1	3
lie	<i>mentira</i>	2	10	2	11	13	1	1
to take off	<i>quitarse</i>	2	10	2	3	4	2	-
goggles	<i>gafas</i>	2	11	3	3	3	1	6
to volunteer	<i>hacer de voluntario</i>	2	11	3	6	6	2	2
waste of time	<i>pérdida de tiempo</i>	2	11	3	4	6	4	-
court	<i>pista</i>	2	12	4	3	3	1	1
to give up	<i>rendirse</i>	2	12	4	2	3	2	-
to worry	<i>preocuparse</i>	2	12	4	3	14	2	1
laces	<i>cordones</i>	2	13	5	2	2	1	5
purse	<i>bolso</i>	2	13	5	3	4	1	3
tooth	<i>diente</i>	2	13	5	8	10	1	2
elective	<i>optativa</i>	2	14	6	3	3	1	12
scholarship	<i>beca</i>	2	14	6	2	2	1	5
to switch	<i>cambiar</i>	2	14	6	3	5	2	1
grease	<i>grasa</i>	2	15	7	2	2	1	4
to check out	<i>comprobar</i>	2	15	7	2	8	2	-
to review	<i>repasar</i>	2	15	7	2	2	2	4
renovation	<i>reforma</i>	2	16	8	2	2	1	5
teen	<i>adolescente</i>	2	16	8	3	3	1	3
termite	<i>termita</i>	2	16	8	4	4	1	14
embarrassed	<i>avergonzado</i>	3	17	1	2	4	3	2
to chill	<i>relajarse</i>	3	17	1	2	3	2	4

Distractors	Translation	Term	Session	Session in term	Freq. in target episode	Total freq. in term	PoS	BNC / COCA word-list
to smell	<i>oler</i>	3	17	1	2	3	2	2
fryer	<i>freidora</i>	3	18	2	2	2	1	3
laundry	<i>colada</i>	3	18	2	2	2	1	6
sauce	<i>salsa</i>	3	18	2	4	4	1	4
diet	<i>dieta</i>	3	19	3	3	3	1	2
goods	<i>productos</i>	3	19	3	3	3	1	2
lotion	<i>crema</i>	3	19	3	2	3	1	9
chimney	<i>chimenea</i>	3	20	4	2	2	1	3
makeup	<i>maquillaje</i>	3	20	4	2	3	1	13
mole	<i>lunar</i>	3	20	4	2	2	1	5
dumpling	<i>empanadilla</i>	3	21	5	3	4	1	6
jelly	<i>gelatina</i>	3	21	5	2	3	1	5
to run across	<i>cruzar corriendo</i>	3	21	5	2	2	2	-
row	<i>fila</i>	3	22	6	2	2	1	2
stage	<i>escenario</i>	3	22	6	2	3	1	1
wife	<i>esposa</i>	3	22	6	7	9	1	1
jaunty	<i>desenfadado</i>	3	23	7	5	5	3	2
sunglasses	<i>gafas de sol</i>	3	23	7	2	2	1	6
to take down	<i>eliminar</i>	3	23	7	4	4	2	-
backpack	<i>mochila</i>	3	24	8	3	3	1	13
speech	<i>discurso</i>	3	24	8	3	3	1	2
trash	<i>basura</i>	3	24	8	3	3	1	10

**Term:** 1= Term 1; 2 = Term 2; 3 = Term 3

**Session:** number of session within the overall intervention (out of 24)

**Session in term:** number of session within the term, from pre- to post-test (from 1 to 8)

**Freq. in target episode:** number of encounters within the target episode (i.e. when the TI was pre-taught in the focused groups)

**Total freq. in term:** total number of encounters with the TI from pre- to post-test within the term

**PoS:** part of speech; 1 = nouns; 2 = verbs (including phrasal verbs); 3 = adjectives; 4 = others (expressions. multiword units)

**BNC/COCA word list:** word family in the BNC/COCA word list corpus (e.g. 1 = 1.000 family band)

Appendix A. 4. Vocabulary test sample

March 2017

Name and surname: \_\_\_\_\_

Class: 2nE (2)

Test de vocabulari

A continuació escoltaràs 40 paraules en anglès. Te les sabràs totes? Escriu-les en anglès a la primera columna i tradueix-les a català o al castellà. Si saps més d'un significat, escriu-lo. Escoltaràs cada paraula dues vegades.



Les 5 primeres paraules són de pràctica. Bona sort!

	ENGLISH	CATALÀ / CASTELLÀ
A	Lucky	Suerte
B	Garden	Jardín
C	To leave	Anar-se
D	Amazing	Espectacular
E	TO SHOW UP	Ensenyar alguna cosa

	ENGLISH	CATALÀ / CASTELLÀ
1	Picasso	Picasso
2	Hand	Mà
3	TO TRUST	confiar
4	christmas eve	Noche buena
5	wolf	Hobo Lobo
6	TO CUT OF	Tallar alguna cosa
7	TO RUSH	
8	Haircut	Corte de pelo
9	TO MAKE SENSE	tenir sentit
10	bank	Banco
11	fourth floor	quarta planta
12	what? weird what?	¿qué? raro?
13	Doll	muñeca (juguete)
14	TO MESS UP	Desastre
15	Sleepover	fiesta de pijamas (dormir molt)

March2017

	ENGLISH	CATALÀ / CASTELLÀ
16	friendship	Amistad
17	Tags	Etiquetes
18	Envelope	Envolucionar
19	Rock	Roca
20	Middle	Al mitg
21	Crownbeery	Fruita
22	Nerd	Frini, empollón
23	to taste	
24	wall	Muro pared
25	wordless <del>wordless</del> <sup>without words</sup>	
26	Turkeys	Pau
27	Oyster	Almeja
28	trap	Trampa
29	Blinder	Ciego
30	to pick on	Agafar alguna cosa
31	to pet / to vet	Acaricia / cuida / cura
32	firework	Fuegos Artificiales
33	to share	Compartir
34	enough	Suficiente
35	flight Attendent	Atendentes de vuelo
36	Review <sup>w</sup>	Revisar
37	to class	A "Estimar"
38	oven	Forn
39	to gible	
40	fract	



21

15

15

FRESH OFF THE BOAT  
Episode: Huangsgiving

Name and surname: \_\_\_\_\_

Class: 2<sup>o</sup> H (B)

Vocabulary training

Relaciona cada palabra con la imagen que la representa:

**3** CRANBERRY

Tomorrow is my best friend's birthday and I'm preparing a **cranberry** cake.

**4** RAW

You can't eat **raw** meat, it's disgusting! First you have to cook it.

**5** CHARADES

I hate playing **charades** because I'm very bad at imitating.

**1** DEER

When I was a kid, my favorite movie was Bambi, because I love **deer**.

**4** HEN

My cousins have six **hens** at home and they have fresh eggs every day.

**2** OVEN

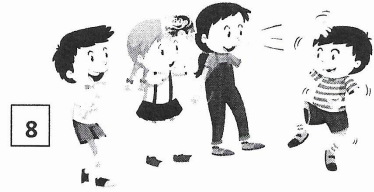
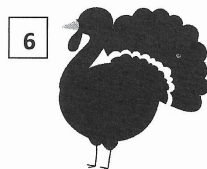
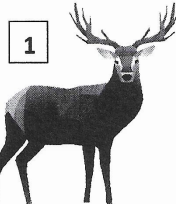
I just put the pizza in the **oven**; it will be ready to eat in 15 minutes!

**5** THANKSGIVING

For **Thanksgiving**, families have lunch together and eat delicious food.

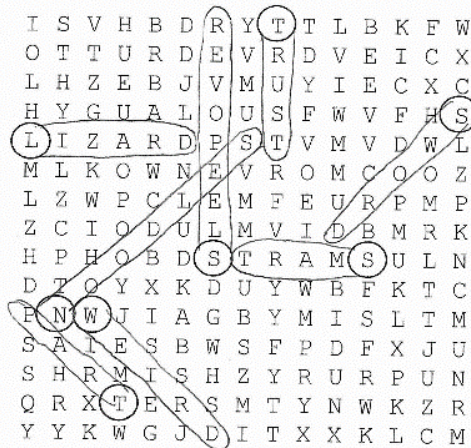
**6** TURKEY

Have you seen the **turkey** there? It has a lot of feathers on its tail!



Vocabulary training

Lee las definiciones y busca las palabras en inglés en la sopa de letras. La primera letra de las palabras está marcada con un círculo. ¡Ánimo!



Definitions

1. Very good at learning or thinking about things (5 letters) SMART ✓
2. A thin, long type of pasta (similar to spaghetti) that is normally eaten in a soup (7 letters) NODDLES ✓
3. Unusual or strange (5 letters) WEIRD ✓
4. A party where one or more people (especially children) stay overnight at one person's house (9 letters) SLEEPOVER ✓
5. A common type of small reptile that has four legs and a long body and tail (6 letters) LIZARD ✓
6. To believe that someone or something is good or honest (5 letters) TRUST ✓
7. A weapon with a long metal blade that has a sharp point, used by knights (5 letters) SWORD ✓
8. A situation in which someone is tricked into doing or saying something (4 letters) TRAP ✓

FRESH OFF THE BOAT  
Episode: The Real Santa

Name and surname: \_\_\_\_\_ Class: 2E(1)

Vocabulary training

Completa los espacios en blanco con la palabra adecuada:

weight	binder	Christmas Eve	boss
makes sense	bakery	fix	enough

1. His height is 1'60 meters and his weight is 55 kilograms.
2. You already ate 3 ice-creams today, I think that is enough!
3. My bike broke down, so I asked my sister to fix it and now looks like new.
4. On Christmas Eve, we sit have supper with our family and eat a lot.
5. If you want to live in New York, it makes sense that you study English.
6. Every Sunday morning my dad goes to the best bakery in town and buys pastries for breakfast for all the family. They are delicious!
7. If you want to have all your notes ordered, you have to put them in a binder.
8. My brother has a new boss who makes them work a lot of hours. He hates him!

DEFINICIONES:

**Bakery:** Tienda donde venden pan y pastas (ensaimadas, madalenas, palmeras...).

**Binder:** Tipo de carpeta con anillas en el centro que sirve para guardar papeles.

**Boss:** Persona que manda por encima de las demás en un negocio.

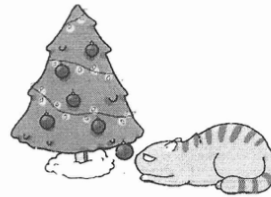
**Christmas Eve:** Noche del 24 de diciembre, el día antes de Navidad.

**Enough:** Lo necesario.

**To fix:** Hacer que algo que estaba mal o estropeado vuelva a estar o ir bien.

**To make sense:** Tener un significado lógico.

**Weight:** La cantidad que una persona o una cosa pesa.



Name and surname: \_\_\_\_\_

Class: 2E(1)

Vocabulary training

Une las palabras con su definición y completa el crucigrama:

HORIZONTAL

- 2. Tener o compartir (algo) con alguien.
- 7. Reunión en la que recibes información o entrenamiento en un ámbito concreto.
- 8. Afecto entre amigos.

VERTICAL

- 1. Plantas con flores de colores brillantes y de formas inusuales.
- 3. Persona que dice tener poderes especiales para predecir lo que va a suceder en el futuro.
- 4. Dejar de tener relación con alguien (amigos, conocidos,...).
- 5. Pieza de ropa u otro material que usan los bebés, que se cambia varias veces al día
- 6. Tira pequeña que se pega a la piel y se usa cuando tienes una herida (como un rasguño o un corte).



Appendix A. 6. Vocabulary post-viewing task sample

**FRESH OFF THE BOAT**

Episode: Showdown at the Golden Saddle

Name and surname: \_\_\_\_\_

Class: ZF(4)

Vocabulary check

Escribe en inglés la palabra que escuches y señala su significado.

- 1) carpool

a) caravana	d) piscina
b) aparcamiento	<input checked="" type="radio"/> c) coche compartido ✓
c) lago	f) No lo sé
- 2) smart

a) tonto	d) avergonzado
<input checked="" type="radio"/> b) inteligente ✓	e) pequeño
c) difícil	f) No lo sé
- 3) to baby sit

a) relajarse	d) hacer de monitor
<input checked="" type="radio"/> b) hacer de canguro ✓	e) dar clases particulares
c) gritar	f) No lo sé
- 4) bulletboard

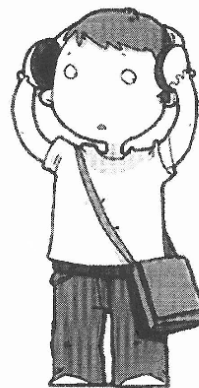
<input checked="" type="radio"/> a) póster ✓	d) papeleo
b) billar	e) pizarra
<input checked="" type="radio"/> c) cartel ✓	f) No lo sé
- 5) franchise

a) francés	<input checked="" type="radio"/> b) franquicia ✓
b) cambio	e) placer
c) local	f) No lo sé
- 6) to steal

<input checked="" type="radio"/> a) robar ✓	d) estar de pie
b) sentir	e) tomar prestado
c) apretar	f) No lo sé
- 7) rack off

a) intestino	d) copia
b) despegue	e) golpe
c) original	f) No lo sé
- 8) buckle

<input checked="" type="radio"/> a) hebilla ✓	d) cinturón
b) burbuja	e) pareja
c) humor	f) No lo sé



1 4 1 3 3 3

Appendix A. 7. Comprehension test sample

**FRESH OFF THE BOAT**  
Episode: The Big 1-2

Name and surname: \_\_\_\_\_

Class: 2G(5)

Comprehension task

**A. Marca la respuesta correcta:**

1. Cuando Jessica le cuenta a Honey que Eddie no quiere una fiesta, Honey...

- a) ...le felicita porque Eddie ya se está haciendo mayor.
- b) ...le advierte que Eddie no está siendo sincero.
- c) ...se sorprende porque Eddie le había dicho que sí que haría una fiesta.



2. ¿Quién ha organizado la fiesta de cumpleaños en el centro comercial?

- a) Mitch
- b) Eddie
- c) Los amigos de Eddie

3. ¿Por qué Evan y Emery empiezan a portarse mal?

- a) Porque quieren ser como Eddie.
- b) Porque sus padres no les hacen caso.
- c) Porque se ha cancelado la fiesta de cumpleaños de Eddie.

3

4. ¿Cuántas horas tarda Jessica en preparar los fideos de cumpleaños?

- a) 3 horas
- b) 5 horas
- c) 6 horas

5. ¿Qué problema tiene Eddie en casa?

- a) Que su madre no le deja invitar a amigos a dormir.
- b) Que hay demasiadas normas.
- c) Que no le dejan comer lo que quiere.

**B. Indica si las siguientes afirmaciones son verdaderas (V) o falsas (F):**

- V /  F A Jessica y Louis no les gusta preparar fiestas de cumpleaños pero las preparan porque se lo piden sus hijos.
- V /  F Jessica convence a Louis para que deje que Eddie vaya a dormir a casa de un amigo.
- V /  F Según Dave, que él y su madre discutan es totalmente normal.
- V /  F Eddie vuelve a casa a medianoche porque le duele la barriga.
- V /  F Jessica acaba castigando a Evan y Emery.

3

Appendix A. 8. Questionnaire BQ1 (pre-intervention)



## Com aprenc millor?

Nom i cognoms: \_\_\_\_\_

Data de naixement: 24/04/2003

Sexe:  Dona  Home

	Mai	Gairebé mai	A vegades	Sovint	Sempre
<b>PART 1. En general...</b>					
1. Recordo millor una cosa si l'escric.	0	1	2	3	4
2. Prenc apunts detallats durant les classes.	0	1	2	3	4
3. Quan escolto, visualitzo imatges, números o paraules mentalment.	0	1	2	3	4
4. Prefereixo aprendre a través de la TV o vídeos en comptes d'altres mètodes.	0	1	2	3	4
5. Utilitzo un codi de colors per ajudar-me a estudiar o treballar.	0	1	2	3	4
6. Necessito instruccions escrites per fer una tasca.	0	1	2	3	4
7. Necessito mirar a la gent per entendre què em diuen.	0	1	2	3	4
8. Entenc millor les classes quan els professors escriuen a la pissarra.	0	1	2	3	4
9. Les gràfiques, diagrames i mapes m'ajuden a entendre millor el que diu algú.	0	1	2	3	4
10. Recordo les cares de la gent però no els noms.	0	1	2	3	4
11. Recordo millor les coses si en parlo amb algú.	0	1	2	3	4
12. Prefereixo aprendre escoltant a un professor més que no pas llegint.	0	1	2	3	4
13. Necessito instruccions orals per fer una tasca.	0	1	2	3	4
14. El so de fons m'ajuda a pensar.	0	1	2	3	4
15. M'agrada escoltar música quan estudio o treballa.	0	1	2	3	4
16. Entenc què diu la gent encara que no la vegi.	0	1	2	3	4
17. Recordo el noms de les persones però no la cara.	0	1	2	3	4
18. Puc recordar fàcilment els acudits que sento.	0	1	2	3	4
19. Puc identificar a la gent per la seva veu (ex. per telèfon).	0	1	2	3	4
20. Quan engego la TV, escolto més el so que no pas miro la pantalla.	0	1	2	3	4
21. Prefereixo començar a fer les coses més que parar atenció a les instruccions.	0	1	2	3	4
22. Necessito pauses freqüents quan treballa o estudio.	0	1	2	3	4
23. Necessito menjar alguna cosa quan llegeixo o estudio.	0	1	2	3	4
24. Si tinc l'opció d'escollir entre seure o estar dret, prefereixo estar dret.	0	1	2	3	4
25. Em poso nerviós/osa quan he de seure quiet/a massa estona.	0	1	2	3	4
26. Penso millor quan em moc (ex. caminant o marcant el ritme amb els peus).	0	1	2	3	4
27. Jugo o mossego els bolígrafs durant les classes.	0	1	2	3	4
28. Manipular objectes m'ajuda a recordar el que diu algú.	0	1	2	3	4
29. Moc les mans quan parlo.	0	1	2	3	4
30. Faig molts dibuixos (gargots) a la llibreta durant les classes.	0	1	2	3	4

**PART 2: Quan miro sèries o pel·lícules en anglès...**

	Mai	Gairebé mai	A vegades	Sovint	Sempre
31. Prefereixo que hi hagi subtítols.	0	1	2	3	4
32. Els subtítols m'ajuden a entendre el diàleg.	0	1	2	3	4
33. Entenc millor l'argument si lleigeixo els subtítols.	0	1	2	3	4
34. Els subtítols m'ajuden a reconèixer les paraules que escolto.	0	1	2	3	4
35. Llegeixo primer els subtítols abans d'escoltar la veu.	0	1	2	3	4
36. Prefereixo veure la sèrie o pel·lícula sense subtítols.	0	1	2	3	4
37. Em concentro més en les imatges que en els subtítols.	0	1	2	3	4
38. Els subtítols em distreuen.	0	1	2	3	4
39. Em concentro més en el diàleg que escolto que en els subtítols.	0	1	2	3	4
40. Em canso llegint els subtítols.	0	1	2	3	4

Amb quina freqüència veus sèries, pel·lícules o programes en versió original en anglès SUBTITULATS...

...en català/castellà?

- mai  
 una o dues hores a l'any  
 una o dues hores al mes  
 una o dues hores a la setmana  
 entre dues i sis hores a la setmana  
 més de sis hores a la setmana

...en anglès?

- mai  
 una o dues hores a l'any  
 una o dues hores al mes  
 una o dues hores a la setmana  
 entre dues i sis hores a la setmana  
 més de sis hores a la setmana

Amb quina freqüència veus sèries, pel·lícules o programes en versió original en anglès SENSE subtítols?

- mai  
 una o dues hores a l'any  
 una o dues hores al mes  
 una o dues hores a la setmana  
 entre dues i sis hores a la setmana  
 més de sis hores a la setmana

Creus que els subtítols t'ajuden a aprendre anglès?

- Gens    No gaire    Una mica    Bastant    Molt

Per què?

Per què escoltes es que diuen i mires els subtítols i relacions sans que signifiquen i aprens paraules noves.

Moltes gràcies!

Appendix A. 9. Questionnaire BQ2 (pre-intervention)

Nom i cognoms

**Aprentatge de l'Anglès**

Benvolguts/des estudiants,

Enguany estem portant a terme un estudi per tal de millorar els exercicis de vocabulari i de *listening* a l'aula, amb la finalitat d'implementar aquest tipus d'activitats regularment a classe. Aquest exercici forma part de l'assignatura de llengua anglesa i per aquest motiu es demana que poseu el nom, però les dades s'anonimitzaran al final per a l'estudi. Tot i així, si algú no desitja que s'inclouin les teves dades ho podràs demanar en finalitzar el quadrimestre.

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			<input checked="" type="checkbox"/>			
Amb subtítols en anglès		<input checked="" type="checkbox"/>				
Sense subtítols			<input checked="" type="checkbox"/>			

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				<input checked="" type="checkbox"/>		
Multijugador		<input checked="" type="checkbox"/>				
MMO (multijugador massiu)	<input checked="" type="checkbox"/>					

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.)						<input checked="" type="checkbox"/>
Llegir llibres, revistes o còmics en anglès			<input checked="" type="checkbox"/>			

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				<input checked="" type="checkbox"/>		
Amb familiars				<input checked="" type="checkbox"/>		
Amb turistes			<input checked="" type="checkbox"/>			
A l'estranger		<input checked="" type="checkbox"/>				

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)			<input checked="" type="checkbox"/>			
Escriure en suport digital (ex: emails, xats, WhatsApp, Facebook, Twitter)				<input checked="" type="checkbox"/>		
Llegir textos (ex: e-books, revistes, pàgines web, blocs, diaris, manuals d'usuari)				<input checked="" type="checkbox"/>		
Mirar vídeos a YouTube						<input checked="" type="checkbox"/>
Escoltar la ràdio / podcasts / música a Spotify						<input checked="" type="checkbox"/>

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					<input checked="" type="checkbox"/>	
Intento segmentar les paraules					<input checked="" type="checkbox"/>	

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		<input checked="" type="checkbox"/>				
Llegeixo els subtítols abans d'escoltar l'àudio				<input checked="" type="checkbox"/>		
Només llegeixo els subtítols si no entenc l'àudio	<input checked="" type="checkbox"/>					

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			X			
Llegeixo els subtítols abans d'escoltar l'àudio			X			
Només llegeixo els subtítols si no entenc l'àudio	X					

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		X				
Utilitzo un traductor online per entendre paraules individuals en anglès				X		

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...

\_\_\_\_\_ X \_\_\_\_\_

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						X
2a estada més llarga						X
3a estada més llarga						X

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.

9 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: perquè s'em dona bé i vull aprofitar

#### E. Informació general

7. Edat: 13 anys

8. Sexe:  Home       Dona

9. Especifiqui la nota obtinguda en l'últim curs d'anglès que vas fer (o a la selectivitat).

Suspens (0 - 4,9)       Aprovat (5 - 6,9)  
 Notable (7 - 8,9)       Excel·lent (9 - 10)

10. 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.)]

No tinc ninguna cosa relacionada

MOLTES GRÀCIES PER PARTICIPAR!



## PART 1 Com aprenc millor en general?

Nom i cognom: \_\_\_\_\_  
 Classe: 2n E  
 Llengua materna: Català i castellà de Catalunya

	Mai	Gairebé mai	A vegades	Sovint	Sempre
1) Quan mires sèries o pel·lícules en anglès...					
1. Prefereixo que hi hagi subtítols.	0	1	2	3	4
2. Els subtítols m'ajuden a entendre el diàleg.	0	1	2	3	4
3. Entenc millor l'argument si lleigeixo els subtítols.	0	1	2	3	4
4. Els subtítols m'ajuden a reconèixer les paraules que escolto.	0	1	2	3	4
5. Lleigeixo primer els subtítols abans d'escoltar la veu.	0	1	2	3	4
6. Prefereixo veure la sèrie o pel·lícula sense subtítols.	0	1	2	3	4
7. Em concentro més en les imatges que en els subtítols.	0	1	2	3	4
8. Els subtítols em distreuen.	0	1	2	3	4
9. Em concentro més en el diàleg que escolto que en els subtítols.	0	1	2	3	4
10. Em canso llegint els subtítols.	0	1	2	3	4

### 2) Quan mires sèries o pel·lícules en versió original en anglès, en general prefereixes...

- subtítols en català o castellà     
  subtítols en anglès     
  sense subtítols     
  no miro mai la versió original en anglès

### Per quins motius mires sèries o pel·lícules en versió original en anglès amb o sense subtítols?

Marca totes les opcions que necessitis a cada columna, i afegeix-ne alguna altra si et fa falta:

Subtítols en català o castellà perquè...	Subtítols en anglès perquè...	Sense subtítols perquè...
<input checked="" type="checkbox"/> M'agrada sentir la versió original <input type="checkbox"/> No hi ha versió doblada <input type="checkbox"/> No hi ha versió subtitulada en anglès <input type="checkbox"/> No m'agraden els subtítols en anglès <input type="checkbox"/> Necessito subtítols per a entendre-ho <input checked="" type="checkbox"/> M'ajuden a entendre-ho millor (però no els necessito) <input type="checkbox"/> Miro les sèries o pel·lícules amb algú que vol/necessita posar subtítols Altres motius: _____ _____ _____	<input type="checkbox"/> M'agrada sentir la versió original <input type="checkbox"/> No hi ha versió doblada <input type="checkbox"/> No hi ha versió subtitulada en cat/cast <input type="checkbox"/> No m'agraden els subtítols en cat/cast <input type="checkbox"/> Necessito subtítols per a entendre-ho <input type="checkbox"/> M'ajuden a entendre-ho millor (però no els necessito) <input type="checkbox"/> Miro les sèries o pel·lícules amb algú que vol/necessita posar subtítols Altres motius: <u>perquè són més fàcils de llegir</u> _____ _____	<input type="checkbox"/> M'agrada sentir la versió original <input type="checkbox"/> No hi ha versió doblada <input type="checkbox"/> No hi ha versió subtitulada <input type="checkbox"/> No m'agraden els subtítols <input type="checkbox"/> No em calen subtítols per entendre-ho <input type="checkbox"/> Els subtítols em distreuen <input type="checkbox"/> Miro les sèries o pel·lícules amb algú que vol/necessita posar subtítols Altres motius: _____ _____ _____



3) Quan mires sèries o pel·lícules en versió en anglès subtítulades en ANGLÈS, què fas primer?

- llegeixo els subtítols       escolto l'àudio       miro la imatge  
 llegeixo i escolto alhora       miro la imatge i escolto alhora       no en miro MAI

I amb subtítulades en la teua LLENGUA MATERNA (cat / cast / etc.), què fas primer?

- llegeixo els subtítols       escolto l'àudio       miro la imatge  
 llegeixo i escolto alhora       miro la imatge i escolto alhora       no en miro MAI

4) Creus que els subtítols t'ajuden a aprendre anglès?

- Gens       No gaire       Una mica       Bastant       Molt

Per què?

Perquè puc veure com s'escriuen les paraules i en quin context puc utilitzar-les.

5) Quines sèries veus habitualment en versió original en anglès?

Game of Thrones

Per què les mires en anglès? Perquè m'ajuda a veure l'original i veure que hi ha canvis i també m'ajuda a saber com pronunciar.

## PART 2

### Sobre l'experiència d'aquest any



1) En general, consideres que has estat atent/a durant els episodis de *Fresh off the Boat*?

- Gens    No gaire    Una mica    Bastant    Molt

2) En general, els episodis de la sèrie t'han agradat?

- Gens    No gaire    Una mica    Bastant    Molt

3) En quins aspectes t'ajuda veure la sèrie? (pots marcar més d'una opció)

- |   |   |
|---|---|
| <input checked="" type="checkbox"/> Connecto paraules escrites amb la seva pronunciació | <input type="checkbox"/> Connecto paraules escrites amb el seu significat |
| <input checked="" type="checkbox"/> Aprenc vocabulari                                   | <input checked="" type="checkbox"/> Sé com s'escriuen les paraules        |
| <input checked="" type="checkbox"/> Milloro el <i>listening</i> en general              | <input checked="" type="checkbox"/> Se'm queda més el que aprenc          |
| <input checked="" type="checkbox"/> Milloro la pronunciació                             | <input type="checkbox"/> Entenc millor la història en general             |

4) De la següent llista, selecciona totes aquelles opcions amb què estiguis d'acord:

- Entenc millor la sèrie ara (els diàlegs, els actors...) que al principi.
- Com a activitat de *listening*, és més natural que altres tipus de *listening*.
- He après aspectes culturals (sobre la cultura xinesa).
- Aquest tipus d'activitat m'ha motivat.
- Hem sentit relaxat/da durant aquesta activitat.
- Estic menys atent que en altres activitats perquè em relaxo a classe.
- Tinc la sensació que he après amb aquesta activitat.
- M'agradaria continuar amb una activitat semblant l'any vinent.

5) Valoració general de l'activitat:

- ☹️    0    1    2    3    4    5    6    😊

6) Cap altre comentari? .....

.....

Thank you!

Appendix A. 11. Questionnaire BQ4 (delayed)

A) Recordes alguna altra paraula que aprenguessis amb la sèrie el curs passat? Quines?

Seer/casor

B) Amb quina freqüència veus ara pel·lícules, sèries o vídeos en anglès a casa?

Subtítols anglès:  mai  1-2h/any  1-2h/mes  1-2h/setmana  2-6h/setmana  +6h/setmana

Subtítols cat/es:  mai  1-2h/any  1-2h/mes  1-2h/setmana  2-6h/setmana  +6h/setmana

Sense subtítols:  mai  1-2h/any  1-2h/mes  1-2h/setmana  2-6h/setmana  +6h/setmana

C) Comparat amb l'any passat veus més pel·lícules, sèries o vídeos en anglès?  si  no  igual

Per què? Perque amb la serie que vam veure l'any passat em va agradar i vaig decidir veure-la.

Prefereixes subtítols en...  anglès  català/castellà  sense  no en miro