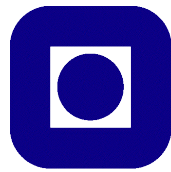


# Effects from using Subtitled Audiovisual Material in Second Language Acquisition

An experimental study in a second language learning classroom in Norway



NTNU  
Norwegian University of  
Science and Technology

Erlend Urkedal York

Master's thesis in English

Department of Language & Literature

Spring 2016



## Abstract

The aim of this master's thesis is to investigate potential long-term effects in second language acquisition from using subtitled audiovisual material as authentic second language input. Over a period of two weeks, 24 Norwegian 16-year-olds watched four episodes of the animated cartoon series Family Guy with English as the source language. The participants were divided into three groups where one group watched with native language (Norwegian) subtitles, one group watched with target language (English) subtitles, and one control group watched without subtitles. Six weeks later, they responded to a word-definition task and a word-recall task. The results indicate that only the group that watched with English subtitles experienced long-term effects. However, the analysis of the results also showed that other significant factors predicted the participants' performance in the post-stimuli tasks, most notably vocabulary proficiency. The approach of the study was quantitative and experimental.



## Preface

The inspiration for this master's thesis was the study "As naturalistic as it gets: subtitles in the English classroom in Norway" by Vulchanova, Aurstad, Kvitnes, & Eschuis (2015). Their study investigated how subtitled audiovisual material can be used to enhance second language acquisition and initially proposed that exposure to such material had both short-term and long-term effects. However, since their study only managed to detect short-term effects from such material, the question about potential long-term effects was left more open. Together with my two supervisors, Mila Dimitrova Vulchanova and Giosuè Baggio, I therefore decided that a study with an increased amount of exposure to audiovisual material would benefit in detecting potential long-term effects.

This project is the culmination of my studies at NTNU, and to my teacher training, before I move on to my future occupation as a teacher in languages. It is therefore appropriate that this master's thesis has investigated how a certain type of input can facilitate second language learning. It is of my opinion that the results in this study can be transferred to several classrooms, especially here in Norway, and that one could use this study as a potential source when deciding how to treat audiovisual material in the second language classroom. Additionally, not only has working with this project allowed me to examine how audiovisual material can be used as authentic second language input, it has also given me the opportunity to learn more about second language acquisition in general.

This project would have been a much greater challenge if not for the help I have received from several people. First, a great thank you to my two advisors, Professor Mila Dimitrova Vulchanova and Associate Professor Giosuè Baggio, who have guided me with invaluable feedback and support. I would also like to thank Sobh Chaboun, who provided sorely needed help with designing the word-recall task. A special thanks to my contact-person from the class used in this study, who not only helped me conduct the experiment, but also came with interesting insights to second language learning and instruction. I wish also to thank fellow student Henrik Eye, who performed a similar study for his master's thesis, for all his technical help and knowledge. An extra warm thanks to my dear sister, Maria Urkedal York, not only for proofreading this thesis, but also for being an inexhaustible source of love and support. Lastly, I wish to thank my partner, Anne

Kari Øhman Meisingset, who has endured my sporadic frustration during my work with the project.

Erlend Urkedal York

Trondheim, May 2016.

## Contents

1. Introduction.....	9
2. Theoretical background .....	13
2.1. Second language acquisition .....	13
2.2. The role of input in second language acquisition.....	15
2.3. Input processing in a L2 .....	17
2.4. Audiovisual authentic material as L2 input.....	19
2.5. Subtitling and dubbing of audiovisual material .....	21
2.6. Subtitled audiovisual material as assistance in SLA .....	23
3. Method .....	27
3.1. Aims of the present study.....	27
3.2. The participants .....	28
3.3. Procedures and research materials .....	30
3.4. Data Analysis .....	33
3.5. Limitations on the study design .....	34
4. Results.....	37
4.1. Vocabulary and grammar tests.....	37
4.2. Word-definition task .....	39
4.3. Word-recall task .....	41
5. Discussion.....	47
5.1. Vocabulary and grammar test .....	47
5.2. Word-definition task .....	47
5.3. Word-recall task.....	50

5.4. General discussion.....	53
6. Conclusion .....	57
Works cited.....	61
Appendices.....	67
Appendix I: Consent form.....	67
Appendix II: The background questionnaire.....	69
Appendix III: The word-definition task .....	77
Appendix IV: The lexical items in the word-recall task .....	89
Appendix V: Boxplots of linear mixed models.....	91



## 1. Introduction

Studies within second language acquisition (SLA) agree that exposure to second language (L2) input is an important part of acquiring L2 proficiency. While input received a secondary role in earlier theories on language acquisition (and consequently SLA), it is now regarded as an essential part of second language acquisition. Accordingly, the perception of how language learners process L2 input has changed over the years, often in relation to developments in current theories on how language is represented and stored in the brain. Recent theories claim that in order for a L2 learner to be able to effectively process and acquire L2 input, one must be provided with comprehensible input and simultaneously given the opportunity to interact with the input. The assumption is that even though comprehension does not necessarily guarantee acquisition, acquisition cannot take place if comprehension does not occur. In other words, processing skills and comprehension are closely related to second language acquisition and in order to acquire language one must acquire processing skills through language interaction.

One way of developing such processing skills is by exposure to authentic second language by means of audiovisual material. This type of L2 input provides learners with authentic language through both an aural and a visual channel. Spoken language and correlating contextual clues observed in the visual material thus offers L2 learners an opportunity to interact with the material and develop skills related to auditory processing and interaction in different ways than other types of L2 input. Consequently, one can claim that authentic language observed in audiovisual material is potentially easier for L2 learners to relate to real-life language interactions, since visual clues and context make it possible to view the message, or language, as much as listen to it, much in the same way as communication takes place in real life. However, although audiovisual material offers a potentially rich source of authentic L2 input, such material must be handled with care if the goal is L2 acquisition. In order to benefit, or acquire, from the audiovisual material, L2 learners must be offered comprehensible L2 input appropriate to their current linguistic level. In other words, exposure to authentic L2 audiovisual material is not always facilitative, and such material must be carefully selected if L2 learners are to benefit from it linguistically.

Several studies have investigated how the use of subtitles can aid learners in making L2 input comprehensible and how different types of subtitles can be used according to linguistic competence. In general, the claim is that interlingual (i.e. from source language audio to native

language text) subtitles will have the most facilitative effect for less skilled learners, and that intralingual (i.e. source language audio and text) subtitles are more beneficial for advanced L2 learners (e.g. Neuman & Koskinen, 1992; Danan, 2004; Mitterer & McQueen, 2009). The claim is that a L2 is able to automatically process native language subtitles, while target language subtitles require more advanced knowledge of the language in order not to interfere with other cognitive processes. In other words, although there is consensus that the use of subtitles can be a valuable tool in SLA, this should also be assessed in the context of the learner's L2 competence.

An original study by Vulchanova, Aurstad, Kvitnes, & Eschuis (2015) investigated potential short-term and long-term effects from subtitles on native Norwegian 16-17 year-old students. In their study, 114 advanced learners of English were exposed to a single viewing of 22 minutes of the American cartoon *Family Guy*. Their study proposed that exposure to subtitles, both interlingual and intralingual, had effects in terms of L2 acquisition; subtitles had short-term effects on comprehension while it did not have long-term effects as measured by a word-recall task.

By increasing stimuli exposure from a single viewing session to four viewing sessions, and from a total amount of 22 minutes to approximately 88 minutes, this study aims to reveal long-term effects from subtitles on a similar, though smaller, group of advanced native Norwegian L2 learners. The aim of this study is thus to investigate the potential long-term effects from using subtitles as an aid in second language acquisition. The main hypothesis is that both interlingual and intralingual subtitles will have an effect on how these advanced learners of English as a second language will perform in post-stimuli tasks. A secondary hypothesis is that these effects are related to different levels of proficiency, and that the participants in the English subtitles group will exceed not only the control group but also the Norwegian subtitles group in terms of post-treatment performance.

24 native Norwegian 16-year-olds from a VGI upper secondary school were exposed to authentic L2 input in the form of audiovisual material. After initial testing of linguistic competence, they had over the course of two weeks four regular viewing sessions where they watched a total of approximately 88 minutes of the American cartoon *Family Guy*. The participants were divided into three equally sized groups, where one group watched with Norwegian (native language) subtitles, one group watched with English (target language) subtitles, and one group

served as the control group and watched without subtitles. The participants completed a word-definition task and a word-recall task six weeks after exposure in order to check for potential long-term effects.



## 2. Theoretical background

### 2.1. Second language acquisition

As the name implies, the second language (L2) of an individual refers to the language acquired after the first language (L1), also known as the native language or mother tongue of an individual. Smith (1995) explains the term “second language” as such: “[it] normally stands as a cover term for any language learned by a given learner or group a) irrespective of the type of learning environment and b) irrespective of the number of other non-native languages possessed by the learner (p.7)”.

Both present and early theories in the field of second language acquisition generally agree that cross-linguistic influences between L1 and L2 play an important role in S2 learning, albeit the view on the amount of influence is widely different (Cook, 2013, p.13). Some researchers claim that these influences are more evident in certain subsystems, such as syntax and morphology, and that it would be possible to estimate whether this influence is beneficial to or leads to difficulty for the learner when facing each subsystem in the process of acquisition of the L2 (Odlin, 2003, p.439). Following this, it would therefore be possible to estimate the total contribution, or effect, of linguistic transfer in an individual learner’s second language acquisition (SLA) (ibid). Other researchers have pointed out that cross-linguistic influences appear in a wide array of different ways and it therefore would be risky to make generalizations about cross-linguistic transfer to a certain language structure, to a subsystem, or to a language as a whole, and that it would be difficult to properly estimate either positive or negative transfer between an L1 and a L2 (ibid. p.442 & Cook, 2008, p.13). Nevertheless, whichever way one regards the predictability of cross-linguistic transfer, recent theories on SLA seem at least to agree that the inescapable difference between L2 acquisition and L1 acquisition is that a major component of the initial state of L2 learning is prior knowledge of a first language (Cook, 2008, p.13.).

Another highly debated topic within SLA research is the relationship between age and L2 competence. The general assumption is that the younger one is at the onset of second language learning, the higher the probability of reaching a high level of competence in the L2. This notion is for instance stated in the critical period hypothesis, which claims that it is only between the age

of two and early teens that humans are capable of learning language, i.e. it is only in this period normal acquisition is possible (Lenneberg, 1967, referenced in Cook, 2008, p.147). Although this hypothesis has been dismissed in later years, evidence still suggest some correlation between age and proficiency in L2. In general, young learners tend to get to a higher level of proficiency in the long term than those who start L2 learning when older, perhaps because the acquisition process slows down with adults (Cook, 2008, p.150). Accordingly, Saville-Troike (2006) suggest that the reason for this is that younger learners are probably more successful in informal and naturalistic L2 learning contexts, while older learners are more successful in formal instructional settings (p.84). In other words, starting young seems to have some advantages, although it is not entirely clear what the reasons for these advantages are (Cook, 2008, p.149).

It has been claimed that there are as many different ways of learning a second language as there are second-language learners. As de Bot, Lowie, & Verspoor (2005) explain, second-language learners may start learning their second language at different ages, may have learned additional languages, may be more or less motivated for learning, may be more or less intelligent, and may have more or less of an aptitude for language learning (p.65). These factors and a multitude of others are important to consider when addressing second-language acquisition and it would be impossible to come to exact conclusions about the effect of any of these without considering the relation between them (ibid. p.65). However, although the different ways of individual SLA are many, it is still possible to make general assumptions on the different processes that takes place during SLA, and theories about the subject have changed over the years. We can roughly divide these theories between an internal and an external focus on the acquisition process.

Based on Noam Chomsky's theory of Universal Grammar (UG), linguistic theories on both first- and second-language acquisition have since the 1960s been centered around the idea that humans are equipped from nature with an innate, cognitive device for language acquisition, a so-called Language Acquisition Device (LAD) (de Bot et al, 2005, p.29). This device contains all the principles that all languages have in common, and in order for a child, or learner, to acquire a language it only needs access to samples from a natural language in order to activate the device (ibid. p.29). Theorists who support Chomsky's idea of an internal language faculty are more focused on the internal factors for language acquisition while theorists who are more focused on a functional approach to language acquisition are more concerned with external factors of language

acquisition. These factors include the use of language in real situations, how the development of linguistic knowledge in L1 or L2 requires communicative use, and how language is used in discourse structures through interaction (Saville-Troike, 2006, p.53).

## 2.2. The role of input in second language acquisition

As Gass & Mackey (2007) eloquently puts it: “Input is the *sine qua non* of acquisition” (p.177). What this means is simply that in order to acquire a language - that being a first-language, a second-language (or third, or fourth, etc.) - one has to be exposed to input from the specific language(s). This input refers to all the language the learner is exposed to, i.e., from listening or reading, or from visual language in the case of sign language (ibid. p.177). Whether based on the notion of innateness, i.e. Chomsky’s language faculty, or on the functional principles of communicative interaction, theories on language learning all stress the importance of language input in order to successfully acquire a language (de Boot et al., 2005, p.51.). As Gass (2003) points out, the perception on how language learners process input and how it interacts with the mental capacities of the learner has changed over the years, both within language studies in general and within SLA research especially (p.229).

The behaviorist focus within psychological and linguistic research was evident within SLA theories in the early part of the twentieth century. L1 acquisition (and consequently SLA) was seen as a child’s ability to imitate and form language habits from the input, and input was therefore considered the principal factor in language acquisition (Gass, 2003, p. 228-229). Later theories gave language input a secondary role in SLA, and followers of the UG approach have traditionally considered acquisition a task that involves children’s induction of a system of rules from the input, guided by the child’s (or learner’s) UG, which all humans are equipped with from nature (Saville-Troike, 2006, p.47). This means that since all L2 learners already have made certain distinctions during their acquisition of a L1 - i.e.acquired a system of rules from parameter settings within the language - one will be able to do the same in the L2 when given language specific input. Saville-Troike (2006) lists possibilities why - based on the role of input within the UG framework - some learners are more successful than others in SLA. These involve (inter alia): “- All learners may not have the same degree of access to UG. – Some learners may receive qualitatively different L2 input from others. – Some learners may be more perceptive than others of mismatches between L2

input and existing parameter settings.” (p.52). However, as Saville-Troike further explains, this approach is mostly focused on the learner’s internal, mental organization of input, and does not adequately explain how language is used in interaction, and how development of linguistic knowledge requires communicative use (ibid. p.52-53).

The term emergentism refers to the research within SLA that share the view that L2 learners use general learning mechanisms in order to extract structures and patterns from the language input they are exposed to (Mitchell, Myles, & Marsden, 2013, p.99). The main idea is that rather than being either innate or learned as abstract structures, grammatical rules and other formal aspects of language “emerge” from language use and experience, such as language specific input (ibid.). In other words, where UG theory is mainly concerned with input as a help to gain access to the universal language found within every human’s mind, the theories related to emergentism identify the language learner as a social being who learns language through cognitive processes where input with different features plays an essential part. Accordingly, since L2 learning necessarily starts after L1 learning, the learner at the onset of his L2 will already have been exposed to a variety of language specific input in the L1, and the different features in the L2 input will therefore affect L2 learning in a different way than L1 input affects L1 learning. The main difference is that the learner is at a developmental stage where he is more socially, interactional, and cognitively mature (ibid. p.98-99). Research on SLA has therefore investigated how particular linguistic features are acquired when exposed to L2 input, i.e. if it is possible to predict how certain characteristics in the input will be acquired late or early. These effects are related to frequency effects from the L2, such as how often a linguistic item appears in the input, and how L2 learners process this (see e.g. Tarone, 2002, p.291).

Following the ideas within emergentism, Stephen Krashen’s (1982) Input Hypothesis formulated the relationship between SLA and the importance of comprehensible input. Krashen claims that in order to acquire language, one must move from one (the current) stage =  $i$ , to the next =  $i + 1$  (p.20). He explains that “a necessary (but not sufficient) condition to move from stage  $i$  to stage  $i + 1$  is that the acquirer understands input that contains  $i + 1$ , where “understands” means that the acquirer is focused on the meaning and not the form of the message” (ibid., p.21). In order for input to be useful for language acquisition, it must therefore be “one step above” from where the learner is at the current moment. However, as Krashen further explains, input should not



deliberately aim at  $i + 1$ , because  $i + 1$  will be provided if the communication is successful (ibid. p.22). The reason why we are able to understand language above our current level is that we are able to use our extra-linguistic information, our knowledge of the context and of the world, to decipher the input directed at us (ibid. p.21).

Although Krashen's Input Hypothesis was a major step towards explaining the link between L2 acquisition and input, some researchers claimed it did not adequately cover the way non-native speakers interact with and how they are able to understand L2 input. Consequently, research started to focus more on the ways the structure of interaction can be modified in order for the input to be more comprehensible for the non-native speaker (Mackey & Abbuhl, 2005, p.208). Especially one type of interaction, and how it affects L2 comprehension and development, has received a lot of attention within SLA research, namely negotiation for meaning. The idea is that learners and their interlocutors make certain efforts in their interaction in order to overcome or avoid difficulties in input comprehensibility (ibid.). Research has therefore studied the relation, and difference, between input simplification and interactionally modified input and their effects on SLA and comprehension (e.g. Ellis & He, 1999, referenced in Mackey & Abbuhl, 2005, p.209). The general distinction between the two types is that the former term refers to input that has been premodified before exposure - e.g. by using more frequent words and less complex grammatical structures -, while the latter refers to input modified through the interaction that takes place after exposure - e.g. when learners ask a speaker with a higher level of linguistic competence, such as a native speaker, when and if they have difficulties in comprehension (ibid.). In other words, the difference between them is related to the timing of the interaction and how this affects the learners' ability to comprehend the L2 input.

### **2.3. Input processing in a L2**

In his Input Processing (IP) theory, VanPatten (2007) aims to illustrate what happens during the processes of language comprehension and the relationship between input and production (output), because, as he explains it: "although comprehension cannot guarantee acquisition, acquisition cannot happen if comprehension does not occur" (p.115). As Tschirner (2011) explains, it is not input in itself that drives SLA, but it is rather a precondition for SLA (p.34). Pienemann (1998) supports this notion, and he proposes that a set of processing procedures function as the basic

condition for SLA, which means that language learners must acquire processing skills in order to acquire language (p.39). Adapted to L2 learning, IP theory is described as a model of the processes involved when learners process L2 input and convert the input into a developing linguistic system, which in turn makes the learners able to create L2 output: INPUT → INTAKE → DEVELOPING SYSTEM → OUTPUT (Lee & Benati, 2007, p.2). However, as Lee & Benati point out, learners filter input, which means that not all input turns into intake, not all intake matches the input, and not all intake is delivered to the developing system (ibid.).

Pienemann (2007) suggests in his Processability Theory (PT) that learners are able to produce and comprehend only those L2 linguistic forms that the current state of the language processor can handle (p.137). Similar to Krashen's Input Hypothesis, PT is based on the assumption that language development is constrained by processability, and in order to move to the next step in development one must be able to process and understand the linguistic items presented in the input. To be able to process effectively a sentence, or an utterance, learners tend to direct their attention towards content words (e.g. nouns and verbs). VanPatten (2007) explains this as The Primacy of Content Words Principle where learners process content words in the input before anything else, and that such linguistic items are less challenging for language learners to process already from an early developmental stage (compared to e.g. morphosyntactic elements in the input) (p.117). Related to this, several studies (e.g. Lee, 2002; White, 1998) claim that input frequency affects both comprehension and word-form recognition in a L2, and that an increased amount of a specific input will greatly enhance input processing and facilitate acquisition. Ellis (2002) supports this claim and further emphasizes the difference between frequency character in the input, such as token frequency and type frequency, and how this plays a part in "[...] processing of phonology, phonotactics, reading, spelling, lexis, morphosyntax, formulaic language, language comprehension, grammaticality, sentence production, and syntax" (p.143). As Ellis further explains (2002), in order for language acquisition to take place, the learner needs to be exposed to these different linguistic items frequently and repeatedly, and perhaps hundreds and hundreds of occurrences of the same items and the same features (ibid. p.144). If the goal is oral proficiency, these items, combinations of items and combinations of items and features need to be experienced aurally. This requires enormous amounts of listening experiences in communicative contexts where the goal, and consequently the result, is comprehension (ibid.)

Related to frequency and the notion of content words is the aspect of salience, i.e. the perceived strength of the stimuli. As Ellis & Collins (2009) point out, input with low salience is less readily learned, and tends to be more difficult for learners to process (p.331). Put simply, one of the main reasons why some items are easier to acquire than others is that they are more prominent in the input than other items. Accordingly, grammatical particles and inflections are of low salience in the language stream whereas lexical units, such as verbs and adjectives (i.e. content words) are stronger psychosocial units, i.e. more related to the social environment of the learner, and thus easier to notice and eventually acquire (ibid.).

#### **2.4. Audiovisual authentic material as L2 input**

Authentic language input is specified as any input material that is not specifically produced for the purpose of second language learning, but rather “[material] that was created to fulfil some social purpose in the language community in which it was produced” (Little, Devitt & Singleton, 1989, p.25). It is now a widely held conviction within SLA research that the use of such authentic material offers L2 learners an advantage in terms of acquisition and that “[it] fairly obviously has a role in fostering contact with and interest in the culture of the target language, and, if sensitively chosen, in making the learning experience enjoyable” (ibid. p.20). In other words, although the language in authentic material is not constructed with the goal of L2 learning in mind, this type of input is still widely recognized as a potent tool for second language learning for both students and instructors.

A predominant source of authentic language input for L2 learners is today found in the form of audiovisual material. As Herron, Morris, Secules and Curtis (1995) point out, film, television, video, and digitized images expose students to large amounts of authentic oral second language input, which in turn can improve L2 linguistic skills such as listening comprehension (referenced in Danan, 2004, p.3). On this list of different types of audiovisual material we can also include video/computer games (see e.g. Sundqvist & Wikström, 2015; Ang & Zaphiris, 2007). Furthermore, Baltova (1994) emphasizes that audiovisual material, since it offers such a rich linguistic context, functions as an effective instructional tool with motivational, attentional, and affective impact on the viewers, which in turn facilitates auditory processing (referenced in Danan, 2004, p.3).

Audiovisual material exposes L2 learners to authentic language/input both in respect of spoken language and (usually) correlating contextual clues observed in the visual material. The combination of sound, picture, and text used in an authentic setting offers L2 learners an opportunity to interact with the material and develop skills related to auditory processing and interaction. As Baltova (1994) explains, authentic audiovisual material is closer to real life because visual clues and context make it possible to “view” the message as much as listen to it (referenced in Danan, 2004, p.3.). Accordingly, in a study of 53 intermediate-level grade 8 Canadian pupils in a core French program, Baltova (1994) found that participants who watched a 15-minute clip in a video-and-sound condition obtained scores almost twice as good as the participants who was exposed to a sound-only condition (ibid.).

Although research within SLA strongly suggest that audiovisual material provides a great potential for authentic second language input, research also emphasizes that such material must be handled with care if the ultimate goal is to be L2 acquisition. As Baltova (1999) remarks, since audiovisual material often contains fast, intertwined, and unclear speech, it is often difficult to process for the learner (referenced in Pavesi & Parego, 2008, p.218). Studies (e.g. Mitterer and McQueen, 2009) suggest that when exposed to such audiovisual material, learners will potentially find it difficult to parse the target language, both aurally and visually (by e.g. subtitles), if their linguistic competence does not match what is required from the material (ibid.). Following Krashen’s theory on comprehensible input and Pienemann’s processability theory, research suggests that in order for L2 learners to benefit effectively from audiovisual material, it is important that L2 learners are offered audiovisual material appropriate to their current linguistic level (see e.g. Neuman & Koskinen, 1992; Danan; 2004). In other words, exposure to authentic L2 audiovisual material is not necessarily facilitative just on its own merit, and such material must be carefully selected in order for it to be pedagogically effective.

As mentioned in an earlier section, one way of ensuring that L2 learners can process target language input is to provide them with the possibility of interaction with the material (Mackey & Abbuhl, 2005, p.208). This is also related to how L2 learners can benefit from authentic input from audiovisual material. A common assumption within SLA research is that learners benefit best from interaction with native speakers or more competent interlocutors and that authentic audiovisual material offers a great potential for such interaction (ibid.). Additionally, in order to develop oral

proficiency skills in a second language, enormous amounts of listening experiences in communicative contexts is required, such as found in authentic audiovisual material (Ellis, 2002, p.144). If second language learners are given the opportunity to negotiate for meaning from the audiovisual material, language that was not comprehensible before now becomes comprehensible because of the negotiation work and is now ready to be incorporated into the learner's target language repertoire (Mackey & Abbuhl, 2005, p.208). In other words, it is possible to modify the structure of interaction between the speaker and the audiovisual material, and hence make the input more comprehensible for the nonnative speaker. However, this modification must be done according to the needs of the learner. As Baltova (1999) remarks, it is pedagogically healthier to provide L2 learners with authentic audiovisual material/input – particularly when subtitled – than simplifying L2 input to suit the learner's level of proficiency (referenced in Aurstad, 2013, p.19). Several studies (which will be introduced later) support Baltova's claim that authentic audiovisual material supported by subtitles can provide L2 learners with comprehensible input, which in turn will facilitate second language learning.

## **2.5. Subtitling and dubbing of audiovisual material**

When addressing the topic of audiovisual translation, and consequently the use of subtitles and dubbing, it is beneficial to first point out the distinction between interlingual and intralingual translation. The former term refers to translation from a source language (e.g. English) into a target language (e.g. Norwegian), while the latter term refers to translation (or more specific: a change of signs) within the same language (Gottlieb, 2004, p.86). While dubbing is mostly concerned with interlingual translation, subtitling is used both in interlingual and intralingual translation (most commonly in subtitles for deaf and hard-of-hearing audiences) (ibid.). Gottlieb (2004) explains further the difference between dubbing and subtitling in terms of (1) semiotics and (2) in difference in wording: (1) While dubbing is interpreted through the same communicative channel as the source language (would have been), subtitles “crosses over” from speech to writing. In other words, dubbing is rendered through the auditory channel while subtitles are rendered through the visual channel and simultaneously the auditory channel. In effect, subtitles supplement the source language, while dubbing substitutes the source language. (2) Whereas subtitles tend to condense the original dialogue by 20-40% (due to technical and perceptual constraints), dubbing is a

speech-to-speech translation, which in general consists of the same amount of linguistic items as the original material. In addition, subtitling is largely governed by the norms of the written language, which is a consideration dubbing does not need to abide to (ibid. p.86-87).

Several countries consistently choose to use dubbing instead of subtitles when translating English-speaking films/television-shows/documentaries/etc. into the native language. In general, it is the largest linguistic communities that choose to dub films and television programmes, while the smaller linguistic communities choose to use subtitles (Gottlieb, 1998; Luyken, 1991, referenced in Blystad & Maasoe, 2004, p.6). However, as Chaume (2012) points out, the distinction between which countries use dubbing and which countries use subtitles is more blurred; and traditional dubbing countries have become more accustomed to subtitling while subtitling countries have become more used to dubbing (p.7.). In general, arguments for and against using either dubbing or subtitles are concerned with questions related to comprehension of the material, and of fluency in the viewing experience (but also in some countries related to nationalistic concerns) (Cintas, 2009, p.4).

In Norway, standard practice when translating audiovisual material is to use subtitles when translating from source language English into target language Norwegian. However, when the audiovisual material is directed to children and a younger audience, common practice is to use dubbing (Blystad & Maasoe, 2004, p.6). In relation to this, several researchers claim to see a tendency of increase in dubbed material to children. As an example, NRK Super (the Norwegian Broadcasting Corporation's children channel) show up to 75% foreign material where most of it is dubbed into Norwegian speech, regardless of source language (Haug, 2013). With this tendency in mind, it is worth noting that age and amount of L2 input has been identified as important factors in SLA. Researchers therefore argue that by dubbing audiovisual material directed to young viewers instead of using subtitles, a great potential for early L2 input, and consequently L2 learning from an early age, is missed (see e.g. d'Ydewalle & Van de Poel, 1999; Koolstra & Bentjes, 1999).

While one can argue in favor of subtitles as the best solution to audiovisual translation, especially in the context of L2 learning, it has also been claimed that this approach is only the lesser of two evils. As Borrell (2002) points out, a common argument against subtitling is that it demands a lot of conscious awareness from the watcher/reader/listener, leading to difficulties in understanding the material (p.4). However, several studies claim that subtitles do not negatively

influence the understanding of the meaning, or coherence, of the audiovisual material (see e.g. Vanderplank, 1988; Borrell, 2000); nor does it damage the general enjoyment of the film or prove to be too cognitively demanding (Perego, Del Missier, & Bottiroli, 2015). As Cintas & Cruz (2008) explain, research on subtitling shows that people have an ability to develop strategies where one effectively utilizes the subtitle as a tool for understanding the material; and repeated exposure to subtitles helps minimize the potential distraction of the text and maximize the usefulness of the subtitles (p.207). Furthermore, because one processes the same word, or phrase, through two different input channels (the auditory and the visual), the claim is that this will benefit in deciphering the meaning of words one has never heard or seen before, since both spoken and written words are closely linked to what is happening on the screen (ibid.). In other words, the claim is that visual input supports auditory processing, which leads to enhanced acquisition.

## **2.6. Subtitled audiovisual material as assistance in SLA**

As mentioned above, L2 input has been identified as one of the major components in SLA, and simultaneously that exposure to authentic audiovisual material in a foreign language – guided by subtitles – can be a powerful aid L2 learning. Related to this is the discussion of what type of subtitles one should choose. This discussion is largely concerned with level of linguistic competence, as to when interlingual (i.e. from source language audio to native language text) subtitles will have the most facilitative effect, compared to intralingual (i.e. source language audio and text) subtitles. The claim is that the use of intralingual subtitles is more beneficial for learners at a higher level of linguistic competence in their L2 compared to less skilled learners (Neuman & Koskinen, 1992; Danan, 2004; Mitterer & McQueen, 2009). Accordingly, additional studies suggest that interlingual subtitles are more effective than intralingual subtitles in facilitating vocabulary acquisition and boost L2 comprehension for less skilled learners (Baltova, 1999; d'Ydewalle & Pavakanun, 1996, both referenced in Pavesi & Perego, 2008, p.216). Bianchi & Ciabattini (2008) propose that the explanation for this difference is that native language subtitles are automatically processed by the viewer, which leaves less of a strain on the cognitive processes involved in listening and in attention to visual content. Target language subtitles on the other hand require a more advanced knowledge of the language in order not to interfere with other cognitive processes (referenced in Aurstad, 2013, p.22). In other words, although there is consensus that the

use of subtitles can be a valuable tool in SLA, the use of subtitles for learning purposes should also be assessed in the context of the learner's L2 competence.

Similar to Bianchi & Ciabattini's claim, Pavese & Parego (2008) suggest that the use of native language subtitles as translation offers an interlingual comparison, which decreases the cognitive effort in language processing, in this way enhancing SLA (p.216). Ivarsson & Carrol (1998) also support this effect and emphasize that viewing an interlingual translation of a foreign language consolidates over time the viewers' familiarity with the language, especially if they already have a working knowledge of it (p.35, referenced in Araújo, 2008, p.228). They suggest that the simultaneous exposure to both interlingual subtitles and to foreign language (i.e. English) audio offers some sort of cognitive assistance for learners of the English language, especially if one is at an early or intermediate competence stage (ibid.). This might help explain why, as Pavese & Parego (2008) report, "[in Europe], people who live in subtitling communities tend to be more fluent in English – even without being formally taught the language – than people living in dubbing communities" (p.216).

Although the use of intralingual subtitles was originally intended as an aid for deaf and hard-of-hearing audiences, it has also become implemented as a tool for L2 learning (Ghia, 2012, p.98). As Ghia points out, research suggests that subtitles in the foreign language help learners visualize what they hear and make it easier to locate word boundaries within speech strings, and to establish correspondences between the graphic form of a word and its phonetic realization (ibid. p.99). In other words, the claim is that target language subtitles help bridge the gap between L2 listening skills and reading abilities. Mitterer & McQueen (2009) support this, and additionally suggest in their study of a group of Belgians, fluent in English, that target language subtitles facilitate L2 understanding and speech learning. They emphasize that a critical factor in native language speech processing is guidance, or retuning, by lexical knowledge of words – i.e. exposure to lexically guided speech-sound categories help learners identify and differentiate between words. They claim that the reason why learners find it difficult to understand a foreign language is the unfamiliar mappings between words and sounds in the foreign language (ibid.). Related to this, Bird & Williams (2002) suggest in a study that the bimodal input of information, which intralingual subtitles offer, contributes to its deeper fixation in both short- and long-term memory (referenced in Ghia, 2012, p.98). In other words, these studies suggest that target language subtitles



help learners differentiate between speech strings and map out the linguistic items of the foreign language, thus strengthening the L2 in both short- and long-term memory. Related to this, Neuman & Koskinen (1992) found in their study that when exposed to target language subtitles, the more fluent L2 students learnt more vocabulary through target language subtitles than the less fluent L2 students did. Similar to Krashen's emphasis on comprehensible input, Neuman & Koskinen (1992) therefore argue that the learner's level of competence in the L2 acts as an intervening variable in mediating the effects of learning through comprehensible input (in this case target language subtitles and target language audio) (p.12).

Vulchanova et al.'s study (2015) of native Norwegian 16 to 19 year olds did not find any major differences between the uses of either interlingual or intralingual subtitles. Their study rather suggests that the mere presence of subtitles as an additional source of information contributes to learners' comprehension of the plot and content in animated audio-visual material in their L2; and that both intralingual and interlingual subtitles can aid target language comprehension in advanced and (especially) very advanced L2 learners (p.9). Similar to Baltova (1999), they found that the combination of auditory material in the target language (L2), verbal visual information, and nonverbal visual information in audio-visual material creates a better environment for learning than when only two of the three are available as input channels (Vulchanova et al, 2015, p. 6-7). Furthermore, their study did not find any long-term effects from the presence of subtitles, but only short-term effects on comprehension (ibid.). Additionally, their study emphasizes that other factors, such as vocabulary size, grammar competence, and daily L2 practices - including watching target language subtitles and playing computer games - are significant predictors of performance on comprehension, "consistent with language learning research and the role of exposure to input" (ibid.). Although Vulchanova et al. (2015) did not encounter any significant long-term effects, their study still suggests that additional exposure to more audiovisual material in the target language, guided by subtitles, can facilitate L2 learning and thus lead to long-term effects in comprehension and proficiency (ibid. p.8).



### 3. Method

As mentioned in the introduction, the aim of this study was to investigate the potential long-term effects on second language acquisition from subtitling, and to check for any difference in these long-term effects in different levels of proficiency. This project is based on a study by Vulchanova et al. (2015), in which they found short-term effects on acquisition but not any long-term effects from subtitling. In order to compare results it was therefore crucial to replicate as much as possible the test conditions from their study. The main difference is that this study has an increased amount of exposure to stimuli: their study used approximately 22 minutes of audiovisual stimuli, while this study has used approximately 88 minutes of audiovisual stimuli, which was shown regularly over a period of two weeks.

25 Norwegian 16-year-olds watched four episodes of the animated cartoon series *Family Guy* with English as the source language, over a period of two weeks. One group of nine students watched with native language (Norwegian) subtitles, one group of eight students watched with target language (English) subtitles, and one control group of eight students watched without subtitles. Six weeks later, they responded to a word-definition task and a word-recall task. The approach of the study was quantitative and experimental and the results were collected in Excel and later analyzed in R.

#### 3.1. Aims of the present study

In order to study the potential long-term effects from subtitles on second language acquisition, and the potential difference in effects according to proficiency, a quantitative research method was chosen to be the most appropriate. When studying a large number of participants, a quantitative approach favors the possibility to compare the results between the participants according to a number of variables (Mackey & Gass, 2005, p.137). In this study, the independent variable was the subtitle language of the audiovisual material: one group was exposed to target language subtitles (English), one group to native language subtitles (Norwegian), and one control group to no subtitles. Since the variable in the study was different between the groups, the research can be said to be experimental. In experimental research one or more variables are deliberately manipulated to determine the effect on another variable in a given population (ibid.). Accordingly,

the participants' test performances will be related to the main variable in the stimuli, i.e. the participants' scores will be assessed in relation to the stimuli type. Additional variables in the test material will be presented in section 3.3.

In order to study the effect of the main variable on the participants, the study was structured with a pretest/posttest design. First, information on the linguistic background of the participants was collected - through both a background questionnaire (see appendix 1) and two online tests on grammar and vocabulary. The participants were then exposed to the audiovisual stimuli, before lastly the participants completed a word-definition task and a word-recall task. Since this project is concerned with investigating potential long-term effects from the subtitles, the participants completed the word-recall task and the word-definition task six weeks after the last exposure to the audiovisual stimuli.

### **3.2. The participants**

This project is based on the results from twenty-four participants. Thirty students, aged 16 to 17, from a VGI class (upper secondary school, grade I) were asked to participate in the study. They were all students in a “studiespesialiserings” – class (preparation for further studies). Although all of the students were encouraged to participate, this did not necessarily mean that it would be possible to include all the participants in the final analysis of the results. For instance, if someone reported not being a native Norwegian speaker then he or she would have to be excluded from the study. Twenty-five of the thirty students choose to participate, and they signed a consent form stating that they wished to take part in the study. However, after assessing the information on their linguistic background one of the participants' results had to be excluded due to the participant having aspects in their linguistic background, that would be difficult to account for in the scope of this project. These aspects were elements that potentially could hinder normal language development, such as having difficulties with reading and writing. This meant that even though twenty-five students participated in the project, only the results from twenty-four participants will be presented in later sections and in the analysis.

As mentioned, this project is inspired by a paper by Vulchanova et al (2015), where three groups of participants were exposed to either Norwegian subtitles, English subtitles, or to no

subtitles. This structure is called a between groups design – or an experimental design - and is often used within second language research (Mackey & Gass, 2005, p.146). The same project design was used for this study, since a secondary aim was to compare results. Therefore, the twenty-five participants were divided into three groups, and relatively at random, but with the criteria that the three groups were more or less equal in size and that the groups were the same for the entire length of the project. Accordingly, the approach of this study is experimental and the two groups that watched with either Norwegian or English subtitles were the two experimental groups, while the group watching without subtitles was the control group.

The participants were divided into three groups as such:

**Experimental group 1 - English subtitles:** Eight of the participants watched the four episodes of Family Guy with English (source language) subtitles. Five were female and three were male. All of these participants' results will be used in the study since no one in this group reported any factors in their linguistic background that would invalidate their results.

**Experimental group 2 – Norwegian subtitles:** Nine of the participants watched all four episodes of Family Guy with Norwegian (target language) subtitles. Of these, two were female and seven were male. Eight of these participants' results will be used in the study since one in this group reported as having difficulties with reading and writing, which is a factor that potentially could invalidate the participant's results.

**Control group – no subtitles:** The eight remaining participants watched without subtitles. Five were female and three were male. All of these participants' results will be used in the study since no one in this group reported any factors in their linguistic background that would invalidate their results.

When applying to NSD (Norsk senter for forskningsdata), one of the most important criteria to meet in order to get the project approved was to make sure that the participants' anonymity was secured. To achieve this, their teacher and my contact person in the class gave each participant a code number. They used this number when they filled out the background questionnaire, when they took the vocabulary test and the grammar test, and when they completed the word-definition task and the word-recall task. Of course, when filling out the consent forms they used their proper names. This meant that their teacher had the information about name and

corresponding code number, while I had the information about code number and corresponding results. It will therefore not be possible to match a name with a result, and it will not be possible to match a code number with a name. It was important to inform the participants that the results were only going to be used for the study and not as a tool for performance-assessment for the teacher.

### 3.3. Procedures and research materials

The participants were not given any information about the particular aim of this study but simply asked if they wanted to contribute to a Master project regarding English as a second language. Those who wanted to partake signed a consent form (appendix II) where they agreed to be present for the entire duration of the project but could discontinue if they wanted to at any time. The project was divided into two parts according to the pretest/posttest design. First, the participants completed two tests related to competence in English grammar and vocabulary on their personal computers. Then they watched four episodes of Family Guy over the course of two weeks. Second, one and a half month after watching the final episode they completed a pen and paper word-definition task (appendix III), and a word-recall task (appendix IV) on two computers.

**Part 1:** The participants were instructed to fill out a background questionnaire, and to answer it as accurately as possible. The aim of these questions was to map out the linguistic background of each participant and to identify factors that potentially can be relevant to second language acquisition, such as their self-estimated level of proficiency in English, time per week spent playing computer games, and any adverse diagnostic aspects (e.g. if they have been diagnosed with ADHD). Some of the questions were formatted as Yes/No questions, while other questions asked the participants to answer with a number between 1 to 5, where 5 equaled to “every day” and 1 to “never”. Questions related to self-evaluation of proficiency were structured with four alternatives ranging from “basic” to “fluent”. These answers were treated in the statistics as numbers between 1 to 4, where “basic” corresponds to 1 and “fluent” corresponds to 4. Furthermore, some of the questions asked for additional information, such as to write down additional languages in which they have competence, and if they have lived in an English speaking country for an extended period. Additionally, one question asked the participants to report their

personal choice of subtitles, i.e. if they prefer using Norwegian subtitles, English subtitles, or no subtitles when watching English-speaking films.

After filling out the background questionnaire, the participants used their personal computers and completed a grammar test and a vocabulary test on the Internet. The vocabulary test can be found at <http://vocabulary.ugent.be/> and the grammar test at [http://www.examenglish.com/cpe/CPE\\_grammar.htm](http://www.examenglish.com/cpe/CPE_grammar.htm). These tests were considered as a reliable method of collecting information on competence related to grammar and vocabulary, given the limitations on my knowledge about each individual participant. The vocabulary test consisted of 100 words where some were existing English words while others were made-up nonwords, and each participant answered Yes or No (J and F on the keyboard) depending on if they knew the word or not. After completion, a score between 0-100 estimated the amount of correct/incorrect answers and this number will be used as an account of the vocabulary competence of each participant. The grammar test contained 50 sentences where each sentence had four possible answers but only one was a correct grammatical expression. The result was given as a number of correct answers, ranging from 0 to 50. In order to prevent false results I personally logged their results when each participant finished both their respective tests.

Immediately after they had completed the grammar test the participants were instructed to go to one of three classrooms, depending on which group they were assigned to. They were not told the reason why they were assigned different groups and different classrooms, only that they were to watch some episodes of the show Family Guy. This was the first of four sessions of exposure to the stimuli, and each session lasted approximately 22 minutes, which is the duration of an episode. One of the groups had a session that lasted a bit longer because of technical problems the session before. Nevertheless, each participant watched four episodes, approximately 88 minutes of Family Guy over a period of two weeks. It is important to note that the participants only watched the episodes once, and in accordance with the experimental design, i.e. one group watched with only Norwegian subtitles, one group of participants watched with only English subtitles, and one group watched without subtitles. The reason why the cartoon Family Guy was chosen for this project was mainly that the original study by Vulchanova et al. (2015) used this as stimuli, but also because of its use of standard American and British accents with a fairly easy plot and well-familiar characters. In addition, as pointed out by Rule and Ague (2005) cartoons create

low affective filters, which causes a high degree of motivation, which potentially could enhance the participant's memory (p.548).

**Part 2:** Six weeks after the final exposure to the audiovisual stimuli the participants were instructed to complete a word-definition task and a word-recall task. While the word-definition task only contained words and expressions that had been mentioned in the episodes, the word-recall task also included fillers, i.e. words that did not occur in the episodes. It is important to note that the test items in the two tasks were related to the audio of the episodes and did not necessarily occur in the subtitles.

The word-definition task (see appendix III) was in the multiple-choice format and contained 40 different words and expressions. Of these 40 test items, ten were idiomatic expressions, such as *mind your own business*. Each word or expression contained four alternative answers: one was the correct definition, one definition was intended to be obviously wrong, and two were intended to be less obviously wrong. The reason why the task was in this format was that this is an effective way of decreasing the probability of consistently getting the right answer by sheer chance. In order to be as efficient as possible the task was completed with pen and paper, since it was estimated that the second task (the word-recall task) would demand a lot of time. In relation to VanPatten's (2007, p.117) Primacy of Content Principle, the participants were only asked to define content words, such as verbs, nouns, and adjectives, since, compared to function words, these items are more prominent parts of the input and also easier to define than function words. While some of the words and expressions were intended to be familiar to the participants, others were intended to be more challenging; and with the increased chance that the participants had not encountered them before watching the episodes. The words and expressions were checked for frequency in the Corpus of Contemporary American English (COCA). For instance, the low frequency adjective *frugal*, and the high frequency noun *surgery* were both included in the task, although it is safe to assume that the participants had encountered at least one of the words at an earlier stage before being exposed to the stimuli.

After finishing the word-definition task the participants completed the word-recall task on two computers. The task consisted of 60 different words, where each word was either a noun, a verb, or an adjective (see appendix IV). Of these 60 words, 35 occurred during the episodes and 25 did not occur during the episodes. As in the word-definition task, these words were checked in



the COCA for frequency, which later will be used in R. The test was designed in such a way that when a word appeared on the screen the participants pressed one of two buttons on the keyboard, according to if they believed they had heard the word in the episodes or not. One of the buttons meant YES while the other meant NO. The participants were not given a time limit to complete the task, but simply instructed to use as much time as needed. The test items appeared in a different order for every participant taking the test in order to counter-balance the test, i.e. reduce the chance of the order of test items to influence the results. Additionally, on one of the computer's keyboard Z equaled YES, and M equaled NO; while on the other computer's keyboard Z equaled NO and M equaled YES.

As mentioned, this task was completed on two computers. One of the computers was borrowed from the Department of Language and Literature at NTNU and the other belonged to a fellow Master's student, who was doing a similar project but in another class and with older participants. It was essential that the two computers were of a high standard (i.e. that they were able to operate at a high level of capacity) in order to ensure that the test results were valid. The test was run using E-Prime, which is a program that is able to log results according to the different features of the test items. For the purpose of this study, the results related to reaction time and accuracy, in relation to word frequency, word class, and occurrence will be the most relevant for the later analysis in R.

### **3.4. Data Analysis**

When collecting such a large amount of data it is important that the data is handled in accordance with statistical treatment. As pointed out by Norris, Ross & Schoonen (2015), quantitative methods within L2 research should aim to help researchers make sense of broad patterns from numeric language (and related) data (p.3). In other words, the general goal of L2 research is to say something about larger tendencies of particular groups and how they tend to respond in particular ways to important social and individual phenomena that are related to language learning. In consequence, if one studies enough cases, with consistency in the methodological methods, one should be able to foresee with some confidence the actual strength with which the variables are related; the actual magnitude of effect that a certain treatment has on its recipients; the actual range in which the values or attributes vary in a group or over time for an individual, and so on (ibid.).

In order to achieve this, the use of quantificational statistics is often explained in terms of descriptive statistics and inferential statistics. Descriptive statistics refers to the most obvious features of the collected data, such as for instance average score and standard deviation, and is used to give an overview of the raw data (Mackey & Gass, 2005, p.250). Inferential statistics on the other hand aims to make generalizations from a sample to a larger population (ibid. p.269).

Since the participants in this study are treated as a representative sample of native Norwegian 16-17 year-olds, the ultimate goal was to investigate their results by means of both descriptive statistics and inferential statistics and to analyze potentially meaningful patterns in the data according to the sets of different variables. In order to give a descriptive account of the data, all of the data was coded and logged into two Excel worksheets. One of the worksheets contained all the background information of the participants (i.e. their answers in the background questionnaire, and their results from the vocabulary test and their grammar test) in addition to their results from the word-definition task. The other worksheet contained all the results from the word-recall task, such as reaction time and accuracy (in relation to the features of the word) when exposed to a word. After obtaining a descriptive overview of the different data sets the average scores were treated by means of inferential statistics, i.e. the average score of the data sets were investigated by the use of the statistical software R, which allowed a further exploration of the results. The tests used in R were the Shapiro-Wilk test, ANOVA (ANalysis Of VAriance), and a linear mixed model. The Shapiro-Wilk test was used in order to check for normal distribution in the data sets. ANOVA was used to analyze the variance in the test items in the word-recall task, i.e. if the different features (or variables) in this task had a statistically significant effect on the test scores. The linear mixed model was used in order to investigate and establish a relationship between the tests conducted prior to exposure to the stimuli and the tests conducted post-exposure. Furthermore, the dependent variables in the ANOVA test were frequency, occurrence and word class. The independent variables were accuracy and reaction time. Accordingly, the ANOVA tests were used to study within-subjects factors and between-subjects factors.

### **3.5. Limitations on the study design**

Since the aim of this project was to check for potential long-term effects from subtitling it was crucial that the participants were exposed to a large set of stimuli. Initially, the participants were

supposed to be exposed to approximately six hours of stimuli. However, because the participants were all students in a class, and the exposure took place during school hours, it was only possible to expose the participants to 88 minutes of stimuli. Furthermore, in accordance with a quantitative study approach it would have been beneficial if an even larger set of participants had been investigated in this study. One can argue that 24 participants is not a sufficient amount of participants to call it a valid sample size of a population. However, this study is more concerned with investigating what the data can tell us beyond the statistics, i.e. identifying the general patterns in the data, and is therefore more exploratory in nature than concerned with statistical hypothesis testing.

The assessment of the different participants' linguistic backgrounds exposed that the participants in the EG condition had lower vocabulary proficiency compared to the two other groups. It would therefore have been beneficial for this study to have the three groups counter-balanced according to proficiency, i.e. that the three groups were more equal in terms of vocabulary competence. If this had been the case, then the effects of the different stimuli conditions would potentially have been more recognizable, since the differences in the pre-treatment factors between the groups would probably then not have influenced group performance as much. However, time limitations prevented that the participants could be organized as such, since the post-treatment tests were performed just prior to the first viewing-session.

When planning this project, a secondary goal was to treat the data as a cross-sectional study of native Norwegian 16-19 year olds. This project tested a group of participants aged 16 to 17, while a fellow Master's student did a similar project investigating 18 to 19 year olds. Unfortunately, because of technical difficulties during stimuli exposure, the two projects ended up using different sets of stimuli material as what was first intended. The two projects applied the same amount of exposure (and from the same source, i.e. Family Guy), but the episodes were not identical. However, it will hopefully be possible to compare the results from the two projects in future research.



## 4. Results

As mentioned in the previous chapter, one of the participant's data was excluded from the project because he/she reported having factors in his/her linguistic background that would potentially invalidate the scores. The rest of the participants' data was organized in Excel and used as input into the statistical program R. All of the data input was treated as average scores, which will be presented in later sections. The distribution of the participants' answers in the vocabulary test, the grammar test, the word-definition task, and the word-recall task was checked in the Shapiro-Wilk normality test. Data from the grammar test, the vocabulary test, and the word-definition task was later treated as independent effect predictors on the dependent variable (the stimuli type). The data in the word-recall task was treated as such: Accuracy and Reaction Time (RT) was entered as dependent variables in R, while the features of the test items (word-class, frequency, and occurrence) was entered as independent variables. In order to check the independent variables' potential effect on the dependent variables the data was analyzed using ANOVA (i.e. analysis of variance) tests of within-subjects factors and between-subjects factors. This will later be considered in relation to the stimuli type. Furthermore, a linear mixed model indicated if there was any statistically significant effect from the participant's scores on the vocabulary and grammar test on the word-definition task and the word-recall task.

### 4.1. Vocabulary and grammar tests

Table 1. | **Mean value, p-value, and standard deviation value in the background vocabulary and grammar tests**

	EG	NG	CG
Vocabulary mean.	45.875	56.875	54.750
Grammar mean.	33.500	34.750	32.000
P-value Voc.	0.635	0.312	0.073
P-value Gram.	0.159	0.245	0.467
Stdev Voc.	13.131	6.556	10.767

Stdev Gram.	4.7207	4.743	5.904
W-value Voc	0.942	0.904	0.838
W-value Gram.	0.873	0.892	0.924

EG = English subtitles group, NG = Norwegian subtitles group, CG = Control group, Mean = Mean score, Voc. = Vocabulary test, Gram. = Grammar test, Stdev. = standard deviation.

As seen in table 1, the English subtitles group scored lower on the vocabulary test than the Norwegian subtitles group and the baseline control group. The scores on the grammar test, on the other hand, were more similar between the three groups, with the control group scoring lower than the two other experimental groups. The Norwegian subtitles group scored higher than the two other groups on both the vocabulary test and the grammar test. One should note that the vocabulary test was measured with a value ranging from 0 to 100, while the grammar test was measured with a score ranging from 0 to 50.

Table 2. | **Linear mixed model of vocabulary, grammar and group**

	Estimate stdev	T-value	Pr(> t )
<b>Vocabulary~Grammar</b>	<b>0.706</b>	<b>6.598</b>	<b>1.39e-10 ***</b>
<b>Grammar~Vocabulary</b>	<b>0.145</b>	<b>6.598</b>	<b>1.39e-10 ***</b>
<b>Vocabulary~Group</b>	<b>4.438</b>	<b>6.871</b>	<b>2.59e-11 ***</b>
<b>Grammar~Group</b>	<b>-0.750</b>	<b>-2.437</b>	<b>0.015*</b>

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Estimate stdev = Estimate standard deviation

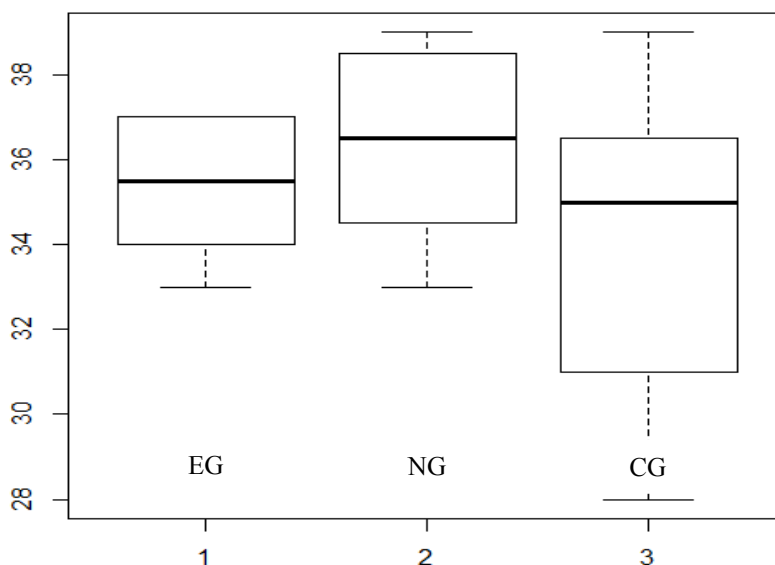
Determined by a linear mixed model, the table above indicates that there was a statistically significant effect between these factors: *Vocabulary* on *Grammar*: ( $T(1,382) = 6.598, p = 1.39e-$

10), *Grammar on Vocabulary* ( $T(1,382) = 6.598, p = 1.39e-10$ ), *Vocabulary on Group* ( $T(1,382) = 6.871, p = 2.59e-11$ ), *Grammar on Group* ( $T(1,382) = -2.437, p = 0.015$ ).

## 4.2. Word-definition task

Figure 1. | **Boxplot of answers in the word-definition task**

---



Note: The word-definition task consisted of 40 questions.

X-axis = Group, Y-axis = Mean score in word-definition task.

EG = English subtitles group, NG = Norwegian subtitles group, CG = Control group.

As seen in Figure 1, the three groups achieved roughly the same mean score in the word-definition task; although the baseline control group scored marginally lower than the two experimental groups. The scores in the word-definition task are similar to the scores in the grammar task where the baseline control group scored lower than the two other groups, and the English subtitles group scored lower than the Norwegian subtitles group. Considering Table 1 and Figure 1, 2, and 3 the Norwegian subtitles group performed better on the vocabulary test, the grammar test, and the word-definition test compared to the two other groups.

Since all the three groups achieved more or less the same score in the word-definition task, it indicates that the choice of subtitle does not predict the outcome of the score in this test, i.e. it suggests that the stimuli type does not have a noticeable effect on the accuracy of the answers in the test. A Shapiro-Wilk normality test shows the distribution of the answers:

Table 2. | **Mean value, p-value, and standard deviation value in the word-definition task – measured by Shapiro-Wilk test.**

	EG	NG	CG
Mean.	35.375	36.000	34.000
Stdev.	1.685	2.264	3.703
P-value	0.080	0.569	0.752
W-value	0.843	0.936	0.954

EG = English subtitles group, NG = Norwegian subtitles group, CG = Control group, Mean = Mean value  
Stdev = Standard deviation.

A linear mixed model indicates a statistically significant relationship/effect on the word-definition task from the participant's results on the vocabulary test and the grammar test:

Table 3. | **Linear mixed model: Vocabulary and grammar effect on the word-definition task**

	Estimate std	T-value	Pr(> t )
<b>Grammar</b>	<b>0.673</b>	<b>7.708</b>	<b>1.12e-13 ***</b>
<b>Vocabulary</b>	<b>1.301</b>	<b>6.627</b>	<b>1.17e-10 ***</b>

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Estimate std. = Estimate standard deviation



The table above indicates a statistically significant relationship/effect between the participants result on the word-definition task and the vocabulary and grammar tasks, as determined by a linear mixed model. *Grammar on Word-definition*:  $T(1,382) = 7.708$ ,  $p = 1.12e-13$ , *Vocabulary on Word-definition*:  $T(1,382) = 6.627$ ,  $p = 1.17e-10$ . See figure 1 and 2 in appendix V for a boxplot of the distribution.

### 4.3. Word-recall task

Table 3. | Mean score, w-value, and p-value of *Accuracy* and *Reaction time* – measured by a Shapiro-Wilk test.

	Acc	RT	W-value Acc	W-value RT	P-value Acc	P-value RT
EG	0.532	1968.638	0.893	0.494	4.048e-08***	2.2e-16***
NG	0.478	1769.584	0.919	0.836	1.154e-06***	1.278e-10***
CG	0.485	2140.358	0.921	0.844	1.43e-06***	2.503e-10***

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Acc = Accuracy, RT = Reaction time.

EG = English subtitles group, NG = Norwegian subtitles group, CG = Control group.

Accuracy was measured with 1 point for a correct answer and 0 for an incorrect answer. Reaction time (RT) was measured in milliseconds. As seen in table 3., the English subtitles group performed better on Accuracy than the two other groups. The baseline control group performed lower than the English subtitles group on both accuracy and reaction time, and lower than the Norwegian subtitles group on reaction time.

Table 4. | ANOVA results (analysis of variance) of within-subject factors.

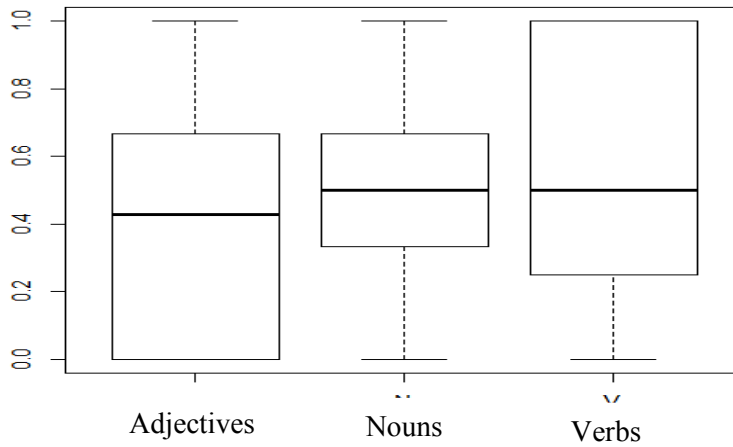
	F-value	Pr(>F)
Acc~Freq	F(2,46)= 0.326	0.723
<b>Acc~Wclass</b>	<b>F(2,46)= 5.205</b>	<b>0.009**</b>
Acc~Occurr	F(1,23)= 3.461	0.076
RT~Freq	F(2,46)= 1.896	0.162
RT~Wclass	F(2,46)= 0.391	0.679
RT~Occur	F(1,23)= 1.77	0.196

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Acc = Accuracy, Freq = Frequency, Wclass = Word class, Occurr = Occurrence, RT = Reaction time.

As we see from the above table, there were no statistically significant differences between subject means as determined by a one-way ANOVA of these variables: *Frequency on Accuracy* ( $F(2,46) = 0.326, p = 0.073$ ), *Occurrence on Accuracy* ( $F(1,23) = 3.461, p = 0.076$ ), *Frequency on Reaction Time* ( $F(2,46) = 1.896, p = 0.162$ ), *Word class on Reaction Time* ( $F(2,46) = 0.391, p = 0.679$ ), and *Occurrence on Reaction Time* ( $F(1,23) = 1.770, p = 0.196$ ). However, there were statistically significant differences between subject means as determined by a one-way ANOVA of *Word class on Accuracy* ( $F(2,46) = 5.205, p = 0.009$ ). This indicates that when analyzing the variance of within subject factors it is only the effect the independent variable *Word class* has on *Accuracy* that is statistically significant. This is illustrated by the figure beneath.

Figure 2. | **Boxplot of Accuracy ~ Word class.**



Note: X-axis = Word class, Y-axis = Accuracy.

Table 5. | **ANOVA results (analysis of variance) of between-subjects factors**

	F-value	Pr(>F)
<b>Acc~Freq</b>	<b>F(4,375)= 2.587</b>	<b>0.004*</b>
Acc~Wclass	F(4,375)= 0.724	0.576
Acc~Occurr	F(1,378)= 1.192	0.305
RT~Freq	F(4,375)= 1.338	0.255
RT~Wclass	F(4,375)= 0.206	0.935
RT~Occur	F(4,375)= 0.937	0.393

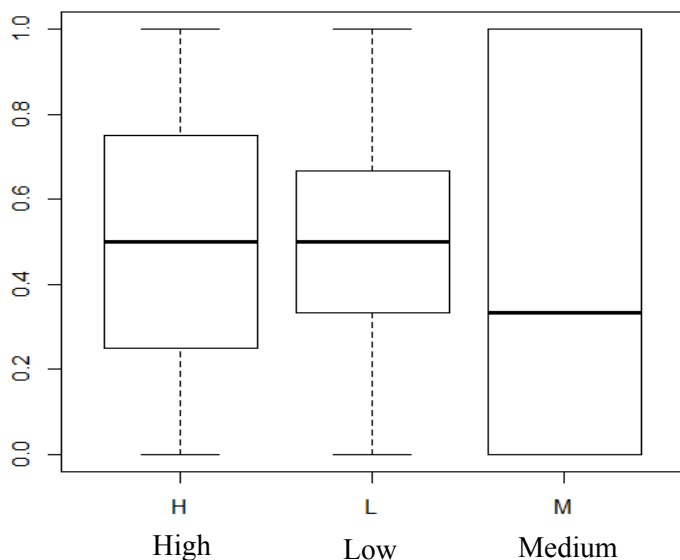
Signif. codes: 0 '\*\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Acc = Accuracy, Freq = Frequency, Wclass = Word class, Occurr = Occurrence, RT = Reaction time.

As illustrated by table 5., there were no statistically significant differences between group means as determined by a one-way ANOVA of these variables: *Word class* on *Accuracy* ( $F(4,375) = 0.724, p = 0.576$ ), *Occurrence* on *Accuracy* ( $F(1,20)=1.192, p = 0.305$ ), *Frequency* on *Reaction*

*Time* ( $F(4,375) = 1.338, p = 0.255$ ), *Word class* on *Reaction Time* ( $F(4,375) = 0.206, p = 0.935$ ), and *Occurrence* on *Reaction Time* ( $F(4,375) = 0.937, p = 0.393$ ). However, there were statistically significant differences between group means as determined by a one-way ANOVA of *Frequency* on *Accuracy* ( $F(4,375) = 2.587, p = 0.004$ ). This suggests that when analyzing the variance of between subject factors it is only the effect the independent variable *Frequency* has on the dependent variable *Accuracy* that is statistically significant.

Figure 3. | **Boxplot of Accuracy ~ Frequency: Between-subjects factors**



Note: The distinction *High*, *Medium*, and *Low* corresponds to words with entries in the COCA as such:  $\text{High} \geq 10000 > \text{Medium} \geq 5000 > \text{Low} \geq 1$ .

Table 6. | **Linear mixed model: Vocabulary and grammar effect on the word-recall task**

	Estimate Std	T-value	Pr(>  T  )
Accuracy ~ Voc.	- 0.0003	-0.186	0.853
Accuracy ~ Gram.	0.001	0.171	0.864

<b>RT ~ Voc.</b>	<b>-13.937</b>	<b>-2.82</b>	<b>0.0051**</b>
RT ~ Gram.	-1.738	-0.158	0.875

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Estimate std. = Estimate standard deviation.

Gram.= Grammar, Voc. = Vocabulary, RT = Reaction time.

Determined by a linear mixed model, the table above indicates that in the word-recall task and related to *Reaction time*, there is a statistically significant effect from the participants’ results in *Vocabulary*: ( $T(1,382) = -2.82, p = 0.00505$ ). See figure 3. in appendix V for a boxplot of the distribution. Simultaneously, the above table indicates that there were no statistically significant effects between these factors: *Vocabulary* on *Accuracy* ( $T(1,382) = -0.186, p = 0.853$ ), *Grammar* on *Accuracy* ( $T(1,382) = 0.171, p = 0.864$ ), and *Grammar* on *Reaction time* ( $T(1,382) = -0.158, p = 0,0875$ ).



## 5. Discussion

### 5.1. Vocabulary and grammar test

In order to create a baseline proficiency measure, all of the participants were tested in English grammar and vocabulary before exposure to the stimuli. The descriptive statistics show that the participants in the Norwegian subtitles group achieved the best mean score on both tests: Vocabulary test = 56.875, Grammar test = 34.750. The baseline control group's performance was slightly lower: Vocabulary test = 54.750, Grammar test = 32.000. The English subtitles group scored notably lower on the vocabulary task than the two other groups: Vocabulary = 45.875, Grammar = 33.500. These results indicate that the English control group initially had lower vocabulary proficiency compared to the two other groups and that this potentially could affect the outcome of the post-tests after the exposure to the audiovisual stimuli. Accordingly, as indicated in a linear mixed model, presented in table 2, both vocabulary and grammar competence had a statistically significant effect on how the participants scored according to group: *Vocabulary on Group* ( $T(1,382) = 6.871, p = 2.59e-11$ ), *Grammar on Group* ( $T(1,382) = -2.437 p = 0.002$ ).

As has been stated earlier, the facilitative effect from the use of subtitles is contingent with comprehension levels. Where Baltova (1999) and d'Ydewalle & Pavakanun (1996) (both referenced in Pavesi & Perego, 2008, p.216) found that native language subtitles are more beneficial for less skilled language learners, Mitterer & MacQueen (2009), Neuman & Koskinen, (1992), Danan, (2004), all observed that target language subtitles are more beneficial for more advanced second language learners. The results in the vocabulary and grammar tests therefore suggest that comparable to the other conditional groups, the English subtitles group should potentially benefit more from watching the English speaking material with native language subtitles, more so than with English target language subtitles. Accordingly, the Norwegian subtitles group should potentially benefit more from target language subtitles. The results from the two baseline tests therefore will be considered later as effect predictors on the post-stimuli tests.

### 5.2. Word-definition task

The three conditional groups achieved roughly the same mean score on the word-definition task, although the baseline control group scored marginally lower than the two experimental groups. Of

a potential score of 40, the English subtitles group achieved a mean score of 35.375, the Norwegian subtitles group obtained 36.000, while the control group reached 34.000. The scores in the word-definition task are similar to the scores in the grammar task where the baseline control group scored lower than the two other groups, and the English subtitles group scored lower than the Norwegian subtitles group. All the three groups achieved more or less the same mean score in the word-definition task, and the descriptive statistics indicate that the choice of subtitles does not predict the outcome of the score in this test. More precise, the descriptive statistics suggest that stimuli type does not have a noticeable effect on the accuracy of the answers in the word-definition task. It is therefore necessary to investigate potential factors that predicate accuracy in the word-definition task by aid of inferential statistics.

A linear mixed model suggests that there is a statistically significant relationship between the participants' scores on the vocabulary test and the grammar test on the word-definition task. The factor *vocabulary* leaves an effect predictor of  $(T(1,382) = 6.627, p = 1.17e-10)$ , while the factor *grammar* produces an effect predictor of  $(T(1,382) = 7.708, p = 1.12e-13)$ . In other words, the results in the vocabulary test and the grammar test are both predictive of the results in the word-definition task. For a further illustration of the relationship between vocabulary and grammar on the word-definition task, see figure 1. and 2. in appendix V. Without going into more detail, a characteristic with inferential statistics is that if the p-value is under the baseline value of 0.05, i.e.  $0.05 > p$ , it is possible to dismiss the null-hypothesis (i.e. that there is no effect on mean values from a variable), and establish that there is a noticeable statistical effect from a variable onto another. Accordingly, the larger T-value a variable has, the larger the evidence against the null-hypothesis. It is therefore possible to suggest that the relationship between the participants' score on the word-definition task are predicted by how they scored on the pre-stimuli tests.

Related to linguistic competence as a dependent factor in comprehension, VanPatten's (2007) Processability Theory and Krashen's (1982) Input Hypothesis state that L2 learners are only able to produce and comprehend those linguistic forms in the second language input that the current state of the language processor can handle. Naturally then, one can predict that the larger vocabulary one has, and the more proficient one is in grammar, the more second language input one is able to process and comprehend; and that the more L2 input one is able to process the better one will do in the word-definition task. In other words, the participants' processing skills and their



ability to comprehend input are related to their vocabulary size and grammar proficiency, and this predicts how good they are at detecting meanings and definitions of words and expressions and thus how they scored in the task.

Similar to Krashen (1982) and VanPatten (2007), Neuman & Koskinen (1992) note that a learner's level of competence in the second language acts as an intervening variable in mediating the effects of learning through comprehensible input. In other words, a L2 learners' preexisting linguistic knowledge determines how and when SLA takes place. One could therefore expect that the participants in the English viewing condition (EG) would score lower on the word-definition task, proportionate with the result in the vocabulary task, compared with the two other groups. Instead, this group produced almost identical results compared to both the Norwegian subtitles group (NG) and the control group (CG). This could indicate, as proposed in the hypothesis, that since the participants in this study are considered as advanced L2 learners, they have benefited more from subtitles in the target language compared to the two other groups and that this explains why they have seemingly improved their vocabulary over the course of this study. However, at this point in the discussion it is potentially risky to bring forth such a conclusion. Nevertheless, one should note that the participants' proficiency skills in vocabulary and grammar seem to predict how they scored in the word-definition task; and that it is not the different stimuli conditions as such that determined how the participants scored in the task. Furthermore, the EG condition shows signs of improving their vocabulary after four weeks of exposure to the audiovisual material.

As stated in an earlier section, input frequency affects both comprehension and word-form recognition in a second language and that an increased amount of a specific L2 input will greatly enhance the L2 learners' processing of input (Ellis, 2002). One can therefore suggest, firstly, that even though the participants were not familiar with a specific word or expression after watching the four episodes of *Family Guy*, they would still be able to calculate, or know its definition since they already have a certain experience with similar types, or tokens from prior L2 input. Second, the results also indicate that the participants already were familiar with the specific words and expressions (i.e. the tests items) in the task before this project. Third, that exposure to the audiovisual material has provided the participants with sufficient input in order to solve the word-definition task and thus achieve comprehension. Which one of these statements is the most accurate is difficult to say. In retrospect, it is possible to argue that it would have been beneficial

for this study to have the participants complete an additional word-definition task prior to exposure to the stimuli, in order to investigate which of the above statements is the most precise.

### 5.3. Word-recall task

The descriptive statistics show that the English subtitles group (EG) achieved a remarkably better result on the word-recall task than the two other groups. Where the EG condition achieved a mean score of 0.532 on accuracy, the two other groups, the Norwegian subtitles group (NG) and the control group (CG), produced a mean score of 0.478 and 0.485 respectively. Since there is a significant difference between the top performing EG group compared to the two other groups, one can suggest that the different viewing conditions predict the participants' score related to accuracy in the word-recall task, and that the use of English subtitles has helped enhance the participants' test scores in the EG condition. It is also interesting to note that the two subtitled conditions in overall needed less reaction time when presented with an item in the task compared to the baseline control group: EG = 1968.638 ms. (milliseconds), NG = 1769.584 ms, and CG = 2140.358 ms. It is therefore necessary to study closer the inferential statistics related to the word-recall task. If the different variables in the testing material, i.e. the test items, have had an effect on the participant's test scores in the task, it will potentially be possible to investigate how this is predicted from to the different stimuli conditions.

A closer inspection of within-subjects factors, as measured by an analysis of variance (ANOVA), shows that the dependent variable, or feature, *Word class* had an effect on the independent variable *Accuracy*: ( $F(2,46) = 5.205, p = 0.009$ ).  $p < 0.05$  indicates that it is possible to dismiss the null-hypothesis and claim that there is an observable effect from a variable onto another; in this case that the dependent variable, or test item feature, *Word class* had an effect on the participants mean score related to *Accuracy*. In other words, the fact that the test items were either a verb, a noun, or an adjective, had a statistically significant effect on whether the participants answered correct or not. Since this particular effect is observed in within-subjects factors, this is related to how the participants scored individually, and not to how they scored as a group belonging to a specific viewing condition.

The reason why *Word class* seems to be predicative of *Accuracy* can be related to the notion of salience, and how “visible” the input is for the learner. As pointed out by Ellis and Collins (2009), input with low salience takes more effort to learn, and are thus more difficult for L2 learners to process. One can therefore suggest that a lexical item, for instance with the feature *noun*, was processed differently from a lexical item with the feature *adjective*, and thus were differently stored in the participants’ lexical memory. In relation to this, VanPatten’s (2007) Primacy of Content Words Principle states that L2 learners look for content words in the input, and accordingly, that L2 learners have the ability to process such words from early on. Since all of the three word classes used in the task – noun, adjective, and verb – can be said to belong to such a Content Principle, one can argue that it is not necessarily the grammatical feature *Word class* that predicts how the participants scored in this task, but rather how prominent the test item had been in the input. In other words, the relationship between a test item and a specific context in the episodes might be more predictive of how the participants performed than the test item’s grammatical realization. The relationship between the test items and context will be discussed further below.

When analyzing between-subjects factors, the dependent variable *frequency* is shown to have had an effect on the independent variable *accuracy*: ( $F(4,375) = 2.587, p = 0.004$ ).  $p < 0.004$  indicates that one can reject the null-hypothesis and thus claim that the feature *frequency* had an effect on how the different groups performed in relation to *accuracy* in the word-recall task. In other words, the frequency of a test item seems to predict the three groups’ test results. As stated in relation to the word-definition task, input frequency is shown to affect both comprehension and word-form recognition in a second language and an increased amount of specific L2 input will greatly enhance the L2 learners’ processing of such input (Ellis, 2002). This is also related to the different types, or tokens of linguistic items the learner is exposed to in the L2 input. Naturally then, if a word (or a similar type/token of word) is already familiar to the learner then it will be easier to process and thus remember.

The test items in the task were chosen based on being as episode specific as possible. However, one should expect that the participants had been exposed to most, if not all, the test items in other L2 contexts/input before participating in the study. As noted earlier, the general assumption is that the more frequent a word is in different types of L2 input, the larger the

probability is that the L2 learner has been exposed to it before, and consequently the larger the possibility is that the learner has stored it in his/her mental lexicon. One of the main ideas behind the word-recall task was therefore that the participants could use the stories, or contexts, in the episodes to help them remember if they had been exposed to the specific words during the viewing sessions or not. One can therefore assume, and which is supported by the results, that words with the frequency feature *low* would be especially potent in this regard (although it might sound contradictory). Since there was an increased chance that the participants had not been exposed to the words with a *low* frequency feature in earlier L2 input, it would potentially be easier for them to associate these types of items with the different contexts observed in the episodes. As an example, it would be fairly easy to relate the low frequency word, *paraplegic* with the episodes, since one of the main characters is a disabled man in a wheelchair. However, the groups seemed to be just as attentive to words with the feature *high*. This is a more surprising, but the reason might be that the high frequency words were closely related to the contexts in which they were observed and that they therefore were potentially easier to associate with what the participants had observed in the episodes.

If it is not the feature type of a word that is predictive of the results, but rather how a word is associated with a certain context, this might explain why vocabulary competence seems to predict the results in the word-recall task. One can assume that the participants with more advanced vocabulary are familiar with more words, both high frequent ones, medium frequent ones, and low frequent ones. Consequently, they would be able to more effectively identify and decide if a word could potentially belong to a certain context or not. They would therefore spend less time deciding, or remembering, if they had been exposed to a specific word or not. This is supported by a linear mixed model, which suggests that vocabulary proficiency had an effect on reaction time: ( $T(1,382) = -2.82, p = 0.005$ ) (see also appendix V, figure 3.). One could assume that the two groups with highest vocabulary competence, i.e. the NG and CG condition, would score better (i.e. spend less time) on reaction time compared to the group that scored lowest, i.e. the EG condition. However, the results indicate that the two subtitles conditions spent less time deciding if they had been exposed to a word or not. This might indicate that the use of subtitles has had an advantage in terms of recollection, and that the use of subtitles has made it so that the participants spent less time trying to remember if a specific word had been mentioned in the episodes or not.

## 5.4. General discussion

The results in this study indicate that both vocabulary and grammatical competence had an effect on the participants' results in both the word-definition task and the word-recall task. Since the three groups achieved more or less the same overall result in the word-definition task, this could indicate that exposure to subtitles does not seem to predict how well they performed. However, the participants in the EG condition were found initially to have lower proficiency in vocabulary compared to the two other groups, but they still managed to achieve a score similar to the two other conditional groups in the word-definition task. Simultaneously, the participants in the EG condition achieved a remarkably better result compared to the two other groups in the word-recall task. This could indicate that the use of English subtitles not only has “bridged the vocabulary gap” between the three groups, but also enhanced the participants in the EG conditions' L2 vocabulary in such a way that these participants have surpassed those participants who were not exposed to target language subtitles.

One can argue that the participants in this study all belong to a group of equally proficient L2 users, and that this is the reason why the three groups have achieved a similar result in the word-definition task. One could also claim that the three groups' difference in proficiency is relatively small compared to other groups of L2 users (found outside this study). Undeniably, the participants in this study all belong to a group of advanced L2 learners. So, although the English subtitles group perhaps is less advanced in English vocabulary compared to the two other groups, they can still be regarded as advanced L2 learners. It is therefore not surprising that the three groups all achieved a high score in the word-definition task. Nevertheless, based on initial vocabulary proficiency, the participants in the EG condition initially would be expected to score lower than the two other groups, but instead they performed equally well. In view of this, research on advanced L2 learners has indicated that all the participants in this study would potentially benefit more from target language subtitles (see e.g. Neuman & Koskinen, 1992; Danan, 2004; Mitterer & McQueen, 2009) than native language subtitles. One can therefore propose that the exposure to bimodal target language input (i.e. English audio and English subtitles) explains why the participants in the EG condition seem to have “caught up” with the participants in the two other groups, and why this group shows signs of improving their vocabulary proficiency.

In a previous section, it was argued that the different linguistic contexts in the stimuli might have helped the participants recollect if they had been exposed to a word or not. It was simultaneously proposed that increased vocabulary and grammatical competence, i.e. word processing skills, predict how effective they are at this: the higher proficiency in vocabulary and grammar, the easier it would be to decide, or process, if a word potentially belongs to a specific context or not. One could therefore suggest that the reason why the three groups achieved such a good result in the word-definition task is that the stimuli has provided the participants with L2 input in specific linguistic contexts, and that this has helped them acquire both form and meaning of certain lexical items. It is therefore possible to claim that exposure to the audiovisual material has given the participants in this study an opportunity to interact with the L2 input in such a way that it has enhanced their L2 acquisition, irrespective from if they were exposed to subtitles or not. It would have been interesting to have a similar group of participants, who had not been exposed to any stimuli (i.e. not been exposed to any of the episodes used in this study) perform an identical word-definition task, and see how well they would perform. This might perhaps be more relevant in an argument for the use of audiovisual material as authentic L2 input in general, and not specifically to the use of subtitles as such. Nevertheless, the fact that the participants in the EG condition were equally successful in matching form with meaning as the two other groups, even though testing prior to treatment suggested that they would not be able to do so, might indicate that the English subtitles made it easier for these participants to associate a linguistic item with a specific meaning.

As suggested earlier, bimodal input of information - such as exposure to simultaneous target language audio and target language text (i.e. intralingual translation) - contributes to its deeper fixation in both short- and long term memory (Bird & Williams, 2002, referenced in Ghia, 2012, p.98). Related to this, the consensus within SLA studies is that when exposed to such bimodal information input, advanced L2 learners are able to process and learn more vocabulary compared to less fluent L2 learners. Put simply, advanced L2 learners benefit more from target language subtitles than native language subtitles and this explains why the participants in the EG condition surpassed the two other groups in the word-recall task, and managed to match the two other groups in the word-definition task. As Neuman & Koskinen (1992) argue, a learner's level of competence in the L2 acts as an intervening variable in mediating the effects of learning through comprehensible input (p.12). Even though the results in this study suggest that it was linguistic

competence - i.e. vocabulary and grammar competence –, and not the use of subtitles, that was predictive of how the participants scored in the post-treatment tests, it is still possible to suggest that the use of English subtitles has had a positive effect in terms of vocabulary acquisition.

The results in this study indicate that the participants in the EG condition have experienced an advantage in terms of vocabulary processing and acquisition, whereas the two other groups, the NG condition and CG condition, do not show similar signs of proficiency enhancement. This therefore seems like a fitting place to suggest that this study would have benefitted from having all the three conditional groups perform a second vocabulary test, after exposure to the stimuli. In this way, it would have been possible to further investigate if the participants had achieved an improvement in their vocabulary proficiency or not; and if the participants' vocabulary skills had either improved, been unaffected, or declined after exposure to the audiovisual stimuli, depending on which condition they belonged to.

The fact that the post-treatment tasks took place six weeks after exposure might seem like an excessively long time if the goal was to find any effects from the subtitles. Admittedly, only one of two experimental groups achieved noticeably better results in the word-recall task compared to the control group. However, this must be viewed in relation to the fact that all the participants in this study are in general regarded as advanced L2 learners, and that they all would potentially benefit the most from target language subtitles. It is therefore possible to claim that the results in this study indicate that only the participants that were exposed to target language subtitles experienced noticeable long-term effects from the subtitles, and that the reason for this is that the participants in the EG condition were the ones who experienced the most beneficiary treatment.





## 6. Conclusion

This study set out to investigate potential long-term effects from watching subtitled audiovisual material in a second language. 24 native Norwegian 16-year-olds were exposed to stimuli in the form of approximately 88 minutes of authentic L2 input through four viewing sessions and over a period of two weeks. The participants were divided into three equally sized groups, where one group watched with Norwegian (native language) subtitles, one group watched with English (target language) subtitles, and one group functioned as the control group and watched without subtitles. In addition to exposure to the stimuli, the participants completed two tests in the form of a vocabulary test and a word-recall task, in order to map their linguistic competence prior to treatment. Six weeks after exposure, the participants completed a word-definition task and a word-recall task.

Recent theories within the study of second language acquisition agree that second language input plays an integral part in acquisition; and that in order for a learner to acquire a L2 one must be exposed to and interact with the second language in different contexts, and often numerous times. Accordingly, in order for a L2 learner to be able to acquire the L2 input, it must be presented in such a way that the learner has an opportunity to process the L2 input effectively. If a L2 learner is able to do so, then this type of L2 input will be recognized as comprehensible input, which will eventually lead to acquisition. As explained by Stephen Krashen (1982) in his  $i + 1$  Input Hypothesis, the reason why learners are able to acquire proficiency in a L2 is that they are able to identify comprehensible input ( $i$ ) and develop processing skills in such a way that they can process those linguistic items they have not yet acquired ( $1$ ). Consequently, in order for input to be useful for language acquisition, it must be one step ahead from where the learner is at the moment. However, input should not purposely attempt to provide  $i + 1$ , because this will be provided if the communication is successful.

A potentially rich source for communicative and authentic L2 input is found in audiovisual material. This type of material has the potential to expose learners to second language both in respect of aural input (i.e. spoken language) and visual input (i.e. subtitles, and correlating contextual clues). As Baltova (1994) argues, since audiovisual material offers a vibrant linguistic context, this type of material, if used correctly, will function as an effective instructional tool with motivational, attentional, and affective impact on the viewers, which in turn will facilitate L2

acquisition (referenced in Danan, 2004,). However, although studies within SLA strongly suggest that audiovisual material is a potential gold mine for authentic second language input, research also emphasizes that such material must be handled with care if the ultimate goal is L2 acquisition. Following Krashen (1982), it is therefore important that L2 learners are offered audiovisual material appropriate to their current linguistic level if they are to benefit, or acquire, from the audiovisual material. In other words, exposure to authentic L2 audiovisual material is not necessarily facilitative just on its own merit, and such material must be carefully selected in order for it to be pedagogically effective.

Several studies suggest that use of subtitles together with L2 audiovisual material has the potential to enhance second language acquisition, and that the use of subtitles has the potential to make input that was initially not comprehensible for the learner comprehensible. However, the use of subtitles must be viewed in relation to linguistic competence. Several studies propose that intralingual subtitles (i.e. target language subtitles) are more beneficial for learners at a higher level of linguistic competence in their L2 compared to less skilled learners (e.g. Neuman & Koskinen, 1992; Danan, 2004; Mitterer & McQueen, 2009). Simultaneously, studies claim that interlingual subtitles (i.e. native language subtitles) are more effective in facilitating vocabulary acquisition and boost L2 comprehension for less skilled learners compared to more advanced learners (e.g. Baltova, 1999; d'Ydewalle & Pavakanun, 1996, both referenced in Pavesi & Perego, 2008, p.216). In other words, although there is consensus that the use of subtitles can be a valuable tool in SLA, this must also be assessed in the context of the learner's L2 competence.

This study initially proposed that exposure to subtitles, both interlingual and intralingual, had potential effects in terms of L2 acquisition. However, unlike Vulchanova et al.'s (2015) original study, which did not find any significant differences between the two different uses of subtitles, the results in this study indicate that it was only one of the subtitle groups, i.e. the English subtitles group, that experienced any significant effects from the use of additional textual input. Simultaneously Vulchanova et al.'s study (2015) did not find any long-term effects from the subtitles. By increasing stimuli exposure from a single viewing session to four viewing sessions, the results in this study indicate that the English subtitles group has experienced long-term effects from the target language subtitles, as measured by a word-recall task. This is in accordance with SLA studies in general, since the participants in this study belong to a group of advanced L2

learners. The results in this study therefore have two main implications: First, target language subtitles are shown to be more beneficial for the least proficient speakers in this study. Second, only target language subtitles were shown to have an effect on advanced L2 learners in terms of acquisition. Further studies are needed and it would have been interesting to perform an additional study of these participants at a later stage in their continuing journey of becoming even more proficient L2 users.



## Works cited

- Ang, C. S. & Zaphiris, P. (2007). Computer games and language learning. In T. Kid & H. Song (Eds.), *Handbook of research on instructional systems & technology* (pp. 449-462). Hershey, PA: IGI Global.
- Aurstad, L. M. G. (2013) *The Role of Subtitles in Second Language Acquisition – An experimental study in the context of the Norwegian School system* (Master's thesis, NTNU). Trondheim: NTNU, Det humanistiske fakultet, Institutt for moderne fremmedspråk
- Araújo, V. L. S. (2008). The educational use of subtitled films in EFL teaching. In J. D. Cintas (Ed.), *The Didactics of Audiovisual Translation*. (pp.227-238). Amsterdam/Philadelphia: John Benjamins Publishing Company
- Blystad, E. & Maasoe, A. (2004). *The Invisible Text. A report on subtitling for Norwegian television*. Oslo: Institute of Media and Communication. University of Oslo.
- Baltova, I. (1999). Multisensory language teaching in a multidimensional curriculum: The use of authentic bimodal video in core French. *Canadian Modern Language Review*, 56(1), 32-48. doi: 10.3138/cmlr.56.1.31
- Borrell, J (2000). *Subtitling or dubbing? An investigation of the effects from reading subtitles on understanding audiovisual material*. Lund: Lund University
- de Bot, K. & Lowie, W. & Verspoor, M. (2005). *Second Language Acquisition: an advanced resource book*. USA, NY: Routledge.
- Chaume, F. (2012) *Audiovisual Translation: Dubbing*. Manchester, United Kingdom: St. Jerome Publishing.
- Cintas, J. D. & Cruz, M. F. (2008). Using subtitled video materials for foreign language instruction. In J. D. Cintas (Ed.), *The Didactics of Audiovisual Translation*. (pp.201-214). Amsterdam/Philadelphia: John Benjamins Publishing Company
- Cintas, J.D. (2009) Introduction – Audiovisual Translation: An Overview of its potential. In J.D. Cintas (Ed.). *New Trends in Audiovisual Translation*. (pp.1-20). Clevedon, GBR: Channel View Publications. Retrieved from <http://www.ebrary.com>.

- Cook, V. (2008). *Second Language Learning and Language Teaching*. London: Hodder Education.
- Danan, M. (2004). Captioning and subtitling: Undervalued language learning strategies. *Meta: Translators' Journal*, 49(1). 67-77. doi: 10.7202/009021ar
- Ellis, N. (2002). Frequency Effects In Language Processing. A Review with Implications for Theories of Implicit and Explicit Language Acquisition. *Studies in Second Language Acquisition*. 24(02). 143-188. doi: <http://dx.doi.org/10.1017/S0272263102002024>
- Ellis, N. & Collins, L. (2009). Input and Second Language Acquisition: The Roles of Frequency, Form, and Function Introduction to the Special Issue. *The Modern Language Journal*. 93(3). 329-335. doi: 10.1111/j.1540-4781.2009.00893.x
- Gass, S, M. (2003). Input and Interaction. In C. J. Doughty & M.H. Long (Eds.) *The Handbook of Second Language Acquisition*. (pp. 224-255). Malden, MA: Blackwell Publishing Ltd..
- Gass, S, M. & Mackey, A. (2005). *Second Language Research: Methodology and Design*. New Jersey, USA: Lawrence Erlbaum Associate, Inc.
- Gass, S, M. & Mackey, A. (2007). Input, Interaction and Output in Second Language Acquisition. In B. vanPatten & J. Williams (Eds.). *Theories in Second Language Acquisition – An Introduction*. (pp.175-199). London: Lawrence Erlbaum Associates
- Ghia, E. (2012). The acquisition of L2 syntax through audiovisual translation. In A. Serban, & A. Matamala, & J-M. Lavaur (Eds.). *Audiovisual Translation in Close-Up: Practical and Theoretical Approaches*. (pp.95-112). Bern: Peter Lang AG, International Academic Publishers.
- Gottlieb, H. (2004). Language-political implications of subtitling. In P. Orero (Ed.). *Topics in Audiovisual Translation*. (pp.83-100). Philadelphia, PA, USA: John Benjamins Publishing Company.
- Haug, C, S. (2013). – *Du gjør ikke barna en tjeneste ved å velge dubbet film*. Retrieved 27.feb 2016, from <http://www.aftenposten.no/kultur/--Du-gjor-ikke-barna-en-tjeneste-ved-a-velge-dubbet-film-7157296.html>

- Koolstra, C.M. & Beentjes, J. W. J. (1999). Children's Vocabulary Acquisition in a Foreign Language through Watching Subtitled TV Programs at Home. *Educational Technology Research and Development* 47(1), 51–60. doi: 10.1007/bf02299476
- Krashen, S. D. (1982). *Principles and Practice in Second Language Acquisition*. Oxford, UK: Pergamon Press Ltd.
- Lee, J. E. (2002). The impact of initial exposure to Spanish future tense morphology through reading in a second language. In S. Gass, K. Bardovi-Harlig, S. Magnan & J. Walz (Eds.). *Pedagogical Norms for Second And Foreign Language Learning and Teaching*. (pp.119-140). Amsterdam: John Benjamins.
- Little, D., Devitt S. & Singleton, D. (1989). *Learning Foreign Languages from Authentic Texts: Theory and Practice*. Dublin: Authentic language learning resources Ltd.
- Mackey, A., & Abbuhl, R., (2005). Input and Interaction. In C. Sanz (Ed.). *Mind and Context in Adult Second Language Acquisition: Methods, Theory, and Practice*. (pp. 207-233). Washington, DC: Georgetown University Press.
- Mitchell, R., Myles, F. & Marsden, E. (2013). *Second Language Learning Theories*. Abingdon: Routledge.
- Mitterer, H. & McQueen, J. M. (2009). Foreign subtitles help but native-language subtitles harm foreign speech perception. *PLoS ONE*, 4(11), 1-5. doi: 10.1371/journal.pone.0007785
- Neuman, S.B., & Koskinen, P. (1992). Captioned television as 'comprehensible input': Effects of incidental word learning from context for language minority students. *Reading Research Quarterly*, 27(1), 95-106. doi: 10.2307/747835
- Norris, J. M., Ross, S. J. & Schoonen, R. (2015) Improving Second Language Quantitative Research. *Language Learning Special Issue: Currents in Language Learning Series: Improving and Extending Quantitative Reasoning in Second Language Research*, 65(1), 1-8. doi: 10.1111/lang.12110
- Odlin, T. (2003). Cross-Linguistic Influence. In C. J. Doughty & M.H. Long (Eds.) *The Handbook of Second Language Acquisition*. (pp. 436-586). Malden, MA: Blackwell Publishing Ltd.

- Pavesi, M. & Perego, E. (2008). Tailor-made interlingual subtitling as means to enhance second language acquisition. In J. D. Cintas (Ed.), *The Didactics of Audiovisual Translation*. (pp.215-225). Amsterdam/Philadelphia: John Benjamins Publishing Company
- Perego, E., Del Missier, F., & Bottioli, S. (2015) Dubbing versus subtitling in young and older adults: cognitive and evaluative aspects. *Perspectives: Studies in Translatology* 23(1), 1-21. doi: 10.1080/0907676X.2014.912343
- Pienemann, M. (1998). Language processing and second language development: Processability Theory. Amsterdam: John Benjamins.
- Pienemann, M. (2007). Processability Theory. In B. vanPatten & J. Williams (Eds.). *Theories in Second Language Acquisition – An Introduction*. (pp.137-154). London: Lawrence Erlbaum Associates.
- Rule, A. C. & Auge, J. (2005). Using humorous cartoons to teach mineral and rock concepts in sixth grade science class. *Journal of geosciences education*, 53(3), 548-558. Retrieved from <http://serc.carleton.edu/files/nagt/jge/abstracts/rule-v53n5.pdf>
- Saville-Troike, M. (2006). *Introducing second language acquisition*. Cambridge, UK: Cambridge University Press.
- Smith, M, S. (1995). *Second Language Learning: Theoretical Foundations*. New York, USA: Longman Publishing.
- Sundqvist, P. & Wikström, P. (2015). Out-of-school digital gameplay and in-school L2 English vocabulary outcomes. *System*, 51, 65-76. doi:10.1016/j.system.2015.04.001.
- Tarone, E. (2002). Frequency Effects, Noticing, and Creativity: Factors in a Variationist Interlanguage Framework. *Studies in Second Language Acquisition*, 24(2), 287-296. UK:Cambridge University Press.
- Tschirner, E. (2011). Video clips, input processing and language learning. In W.M. Chan, K.N. Chin, & M. Nagami (Eds.). (pp. 25-42.), *Media in Foreign Language Teaching and Learning*. Berlin/Boston, DE: De Gruyter Mouton.



- Vulchanova, M. & Aurstad, L. M. G & Kvitnes, I. E. N & Eschuis, H. (2015). As naturalistic as it gets: subtitles in the English classroom in Norway. In V. Chondrogianni (Ed.), *Frontiers in PSYCHOLOGY*, 5, 1-10. doi: 10.3389/fpsyg.2014.01510
- Vanderplank, R. (1988). The value of teletext sub-titles in language learning. *ELT Journal*, 42(4), 272-281. doi:10.1093/elt/42.4.272
- VanPatten, B. (2007). Input Processing in Adult Second Language Acquisition. In B. VanPatten and J. Williams (Eds.) (pp.115-135). *Theories in Second Language Acquisition*. Mahwah, New Jersey, USA: Lawrence Erlbaum Associates.
- White, J. (1998). Getting the learners attention: A typographical input enhancement study. In C. Doughty and J. Williams (Eds). (pp.85-113). *Focus on Form in Classroom Second Language Acquisition*. Cambridge, MA: Cambridge University Press.



## Appendices

### Appendix I: Consent form

#### Samtykke til deltakelse i undersøkelse om andrespråkforståelse

**Ansvarlig institusjon:** NTNU.

**Student:** Erlend Urkedal York (york@stud.ntnu.no).

**Veiledere:** Mila Vulchanova og Giousuè Baggio.

Vi ønsker å gjennomføre en undersøkelse i din klasse der vi med utgangspunkt i engelsk som andrespråk vil se på din kompetanse i og forståelse av det engelske språket.

Studien vil bestå av to deler og begge deler vil foregå i skoletiden. *Del 1* vil foregå i september/oktober, og *del 2* vil foregå i februar. *Del 1* innebærer at du skal være med på en kort kartlegging av din språklige bakgrunn og kompetanse i engelsk, samt at du over en periode på 4 uker skal se 1 filmklipp per uke. Denne delen inkluderer også en kartlegging av eventuelle diagnoser o.l. som kan være relevant for språklæring. *Del 2* inneholder språklige tester knyttet til filmklippene.

En kode knytter deg til dine opplysninger gjennom en deltakerliste. Det er kun autorisert personell knyttet til prosjektet som har adgang til deltakerlisten og som kan finne tilbake til informasjonen. Det er kun læreren som har oversikt over hvilke navn som er knyttet til koden. Læreren vil ikke ha tilgang til dine resultater i studien, og studien vil derfor ikke kunne ha innvirkning på dine karakterer. Denne oversikten vil også bli slettet når studien er ferdig. All informasjon vil bli anonymisert ved prosjektslutt, og alle opplysninger gitt i undersøkelsen vil bli behandlet konfidensielt. Det vil ikke være mulig å identifisere deg i resultatene av studien når disse publiseres. Skolen vil også være anonymisert.

Selv om du ikke skal delta i studien må du fortsatt være til stede i undervisningen under gjennomføringen av studien. Man får altså ikke fri til å gjøre andre ting mens undersøkelsen pågår.

Deltakelse i undersøkelsen er frivillig, og du kan når som helst trekke deg fra undersøkelsen underveis uten å oppgi en spesiell grunn.

Dersom du vil delta så vennligst fyll ut og lever denne samtykkeerklæringen til din engelsklærer/faglærer.

Vi ber om at skjemaet leveres så raskt som mulig for at du skal kunne delta.

Gjerne ta kontakt med Erlend ved eventuelle spørsmål.

Jeg samtykker til å \_\_\_\_\_ (ditt navn) delta i undersøkelsen.

Trondheim, Dato: \_\_\_\_\_

Underskrift: \_\_\_\_\_.

## Appendix II: The background questionnaire

### Bakgrunnsinformasjon for forskningsprosjekt om andrespråkforståelse

Tusen takk for at du har sagt ja til å delta i vårt forskningsprosjekt om andrespråkforståelse. I dette skjemaet ber vi om bakgrunnsinformasjon som er nødvendig for at resultatene fra undersøkelsen skal kunne brukes.

Alle opplysninger du gir her vil senere bli behandlet uten direkte gjenkjenning opplysninger. En kode knytter deg til dine opplysninger gjennom en deltakerliste. Det er kun autorisert personell knyttet til prosjektet som har adgang til deltakerlisten og som kan finne tilbake til infoen. Del B, C og D av dette skjemaet vil bare oppbevares med koden. All informasjon vil bli anonymisert ved prosjektslutt. Det vil ikke være mulig å identifisere deg i resultatene av studien når disse publiseres.

Legg merke til at skjemaet har 6 sider.

Med takknemlig hilsen,

Erlend Urkedal York / Henrik Eye

Studenter ved lektorutdanningen med master i språk, NTNU.

#### **Del A: Personlig informasjon.**

**Studieretning og trinn:** \_\_\_\_\_.

**Fødselsår:** \_\_\_\_\_.

**Kjønn:**  **Kvinne**  **Mann**

**Bostedskommune:** \_\_\_\_\_.

Deltakerkode:
---------------

**Del B: Språklig bakgrunn:**

- **Morsmål**

Er norsk morsmålet ditt?  Ja  Nei

Hvis ja, har du andre morsmål i tillegg?  Ja  Nei

Hvis ja, hvilke(t) språk? \_\_\_\_\_.

Hvilket språk bruker dere hjemme? \_\_\_\_\_.

**Hvor ofte leser du tekst skrevet på norsk?**

Hver dag  Flere ganger per uke  Et par ganger i uken  c Av og til  Aldri

**Hvor ofte skriker du tekst på norsk?**

Hver dag  Flere ganger per uke  Et par ganger i uken  c Av og til  Aldri

- **Engelsk og andre fremmedspråk**

I engelsk, hvordan vurderer du ferdighetene dine på hvert av disse områdene?

	<b>Grunnleggende</b>	<b>Middels</b>	<b>Avansert</b>	<b>Flytende</b>
<b>Lesing</b>				
<b>Skriving</b>				
<b>Snakke</b>				
<b>Lytte</b>				

**Har du bodd i, eller hatt lengre opphold i et land hvor engelsk er hovedspråk?**

Ja     Nei

Hvis ja, hvor lenge varte oppholdet, oppholdene? \_\_\_\_\_.

**Har du vært på kortere (under 14 dager) reise i et land hvor engelsk er hovedspråk?**

Ja     Nei

**Har du bodd i, eller hatt lengre opphold i et land hvor annet enn engelsk er hovedspråk?**

Ja     Nei

Hvis ja, hvor var det, og hvor lenge varte oppholdet/oppholdene?

**Hvilket språk kan du utover morsmålet ditt og engelsk?**

<b>Språk</b>	<b>Nivå</b>			
	<b>Grunnleggende</b>	<b>Middels</b>	<b>Avansert</b>	<b>Flytende</b>
Tysk				



Fransk				
Spansk				
- angi språk				
- angi språk				
- angi språk				

**Hvor ofte leser du tekster på engelsk?**

Hver dag    Flere ganger per uke    Et par ganger i uken    c Av og ti    Aldri

**Hvor ofte skriver du tekster på engelsk?**

Hver dag    Flere ganger per uke    Et par ganger i uken    c Av og ti    Aldri

**Hvor ofte lytter du til/hører du engelsk?**

Hver dag    Flere ganger per uke    Et par ganger i uken    c Av og ti    Aldri

**Hvor ofte ser du engelskspråklige serier/filmer?**

Hver dag    Flere ganger per uke    Et par ganger i uken    c Av og ti    Aldri

**Når du ser engelskspråklige filmer, hvilke av disse alternativene bruker du oftest?**

Undertekst på norsk (morsmål)    Undertekst på engelsk    Ingen undertekst

**Hvor ofte ser du engelspråklige tegneseriefilmer/serier?**

Hver dag    Flere ganger per uke    Et par ganger i uken    c Av og ti    Aldri

**Har du sett tegneserien «Family Guy»?**

Ja    Nei

Hvis ja, i hvor stor grad? \_\_\_\_\_.

**Hvor ofte spiller du engelspråklige dataspill?**

Hver dag    Flere ganger per uke    Et par ganger i uken    c Av og ti    Aldri

Hvilken type spill spiller du? \_\_\_\_\_.

Hvor mange timer per dag? \_\_\_\_\_.

**Hvor mye TV ser du hver dag?**

7 timer eller mer    5-6 timer    3-4 timer    1-2 timer    c Ser ikke TV

## **Del C: Andre faktorer i språklæring**

**Har du, eller har hatt, problemer med synet utover normal, brillebruk?**

Ja     Nei

**Har du, eller har hatt, problemer med hørselen?**

Ja     Nei

**Har du, eller har hatt, språkvansker av noe slag (spesifikke språkvansker, lese-/lærevansker eller lignende)?**

Ja     Nei

Hvis ja, spesifiser: \_\_\_\_\_.

**Har du, eller har hatt, andre diagnoser som kan tenkes å påvirke språklæring (ADHD, autisme eller lignende)?**

Ja     Nei

**Er du venstrehendt?**

Ja     Nei

**Del D: Vokabulartest og grammatikktest**

**Resultat vokabulartest:**

**Resultat grammatikktest:**

### Appendix III: The word-definition task

Note: Correct answers are marked with X.

Participant code:
-------------------

**Select the most appropriate definition for each word and expression. Select only one alternative for each word. Mark your answer with X.**

**1. Mature.**

- a. A characteristic of food that is made from tomatoes.
- b. A characteristic of a person who is fully developed, often in terms of body and mind. X
- c. A characteristic of a person who is often tired.
- d. A characteristic of a person who is young.

**2. To make up one's mind.**

- a. To conclude a chapter in a book.
- b. To choose the direction of a play.
- c. To paint one's face.
- d. To come to a decision. X

**3. A sundae.**

- a. A drink made from lemons.
- b. The last day of the week.

- c. Ice cream served with syrup over it. X
- d. A person from Italy.

**4. To stumble.**

- a. To fall, trip. X
- b. To shave one's beard.
- c. To have trouble with something.
- d. To kick something.

**5. Swear to God.**

- a. To promise something. X
- b. To say a bad word when praying to God.
- c. To lie about something.
- d. To feel upset about something.

**6. Dry cleaners.**

- a. A place that washes dried fruit.
- b. A person who cleans the streets in a city.
- c. A shop that sells vacuum cleaners.
- d. A business that uses a certain process to clean clothes. X

**7. In a rush.**

- a. To have a lot of adrenaline.
- b. To be in a queue.
- c. To be sad.
- d. To be pressed for time, be in a hurry. X

**8. Frugal.**

- a. A characteristic of someone who is economical, someone who does not use much money and who is not wasteful. X
- b. A characteristic of someone who wears colorful clothes.
- c. A characteristic of someone who goes to bed early in the evening.
- d. A certain type of flower from Japan.

**9. Wit.**

- a. A statement by a person who is seen as less intelligent.
- b. A certain type of intelligence, often related to quick thinking. X
- c. A type of ship that was common during the Middle Ages.
- d. An exceptional piece of literature.

**10. Honeymoon.**

- a. A dessert.
- b. A type of apple.
- c. The day after a full moon.
- d. A vacation or trip taken by a newly married couple. X

**11. A pain in the ass.**

- a. A suppository pill.
- b. Something causing trouble. X
- c. A donkey with problem.
- d. Constipation.

**12. Restroom.**

- a. A room that contains facilities such as toilet, washbowl and sink. X
- b. A room often found in airports with a bed where you can have a quick nap.
- c. A room to keep your pets.
- d. A device used to restring a guitar.

**13. To cheer up.**

- a. To buy new furniture.
- b. To view something from above.
- c. To become happy. X
- d. To sing cheerfully.

**14. A tissue.**

- a. A type of trousers made from denim.
- b. An appendix to a scientific paper.
- c. A piece of thin paper used for purposes such as cleaning your nose. X
- d. A part of a camera.

**15. A novel**

- a. A type of poem written in the 1800s.
- b. A fictional prose narrative of considerable length. X
- c. A prose narrative consisting of minimum two volumes.
- d. An encyclopedia.



**16. Unconscious.**

- a. Without awareness, sensation, cognition. X
- b. Without care for others.
- c. Ambitious.
- d. Have a wish to change something.

**17. To date someone.**

- a. To go out socially with someone else, often in a romantic way. X
- b. To celebrate someone's birthday.
- c. To have a boyfriend/girlfriend.
- d. To find new friends.

**18. A will.**

- a. A short piece of printed-paper on a current topic.
- b. A wish, often in the form of a document called a testament. X
- c. The nickname of a person with the name William.
- d. A flying piece of printed-paper on a current topic.

**19. Apologetic.**

- a. Sorry, regretful. X
- b. Astronaut.
- c. Hopeful.
- d. Patient.

**20. To have a broken heart.**

- a. To experience great emotional pain. X
- b. To have cardiac arrest.
- c. To feel a strong pain in the chest.
- d. To be broke.

**21. A waiter.**

- a. A person who makes food in a restaurant.
- b. A person who is impatient.
- c. A person, especially a man, who serves tables in a restaurant. X
- d. A person who makes drinks in a bar.

**22. A freak accident.**

- a. An accident created by someone who does not look like a human being.
- b. A car crash.
- c. An accident caused by someone with good intentions.
- d. An incident, especially one that is harmful, occurring under highly unusual and unlikely circumstances. X

**23. To figure it out.**

- a. To make a sculpture.
- b. To realize, or understand, how something works. X
- c. To explain how something works.
- d. To throw something in the garbage.

**24. To demonize.**

- a. To pretend to be a monster.
- b. To play a simple version of a game.
- c. To claim that something, or someone, has bad qualities. X
- d. To act as if you are possessed by a demon.

**25. Treadmill.**

- a. An exercise machine that allows the user to walk or run in place. X
- b. A machine that makes yarn from wool.
- c. A machine/building that uses wind-power to grind material.
- d. A hair curler.

**26. “No way”.**

- a. “You are stupid”.
- b. “Not at all”. X
- c. “Not a road”.
- d. “Never mind”.

**27. Committed.**

- a. To be happy about a change.
- b. To leave someone out.
- c. To be bound to something. X
- d. To evolve into something more serious.

**28. Surgery.**

- a. Treatment, such as an operation, performed by a surgeon. X
- b. A brand of soda.
- c. Situation where a person is in the immediate danger of losing a limb.
- d. A factory that processes sugar.

**29. Jukebox.**

- a. A box to store jewelry.
- b. A box that contains all the evil powers in the world.
- c. A person who watches too much television.
- d. A machine that lets you select and play a song. X

**30. Packed.**

- a. Intoxicated; drunk.
- b. Exhausted; tired.
- c. Starving; hungry.
- d. Filled to capacity; full. X

**31. Policy.**

- a. A course of action adopted and pursued by a government, ruler, political party, etc. X
- b. The police department's mascot.
- c. A set of moral values.
- d. A common nickname for a police officer.

**32. Establishment.**

- a. A place of business together with its employees, merchandise, equipment, etc.  
X
- b. A building used for breeding and storing livestock such as cows and oxen.
- c. Superman's headquarters.
- d. The supreme court in the USA.

**33. "Mind your own business".**

- a. "This does not concern you". X
- b. "Take care of your company".
- c. "You need help".
- d. "This is important".

**34. Offspring.**

- a. Children or young of a particular parent. X
- b. A band from the US who become famous for their song Smells Like Teen Spirit.
- c. A reservoir for storing large amounts of water.
- d. Bottled water that has gone past its best-before date.

**35. Dignified.**

- a. Buried under ground.
- b. Characterized or marked by an aspect of dignity. X
- c. An album by Britney Spears.
- d. Scared of darkness.

**36. The gist of something.**

- a. Central idea. X
- b. A protagonist.
- c. A conception.
- d. A poison.

**37. Fellow.**

- a. A person who walks behind others.
- b. A person who works on a farm.
- c. A person who belongs to an organization.
- d. A man or a boy. X

**38. To mooch.**

- a. To borrow an item without intent to return or repay it. X
- b. To pretend you are a cow.
- c. To eat a large amount of food, even though you are not hungry.
- d. To spend a lot of money on an item.

**39. Keep a low profile.**

- a. To try to not attract much attention. X
- b. To try to not grow much in height.
- c. To prefer to buy a certain type of car tires.
- d. To take great care of one's appearances.

**40. To point fingers.**

- a. To blame someone. X
- b. To insult someone.
- c. To admit something.
- d. To have a finger with a sharp nail.

Thanks for participating!





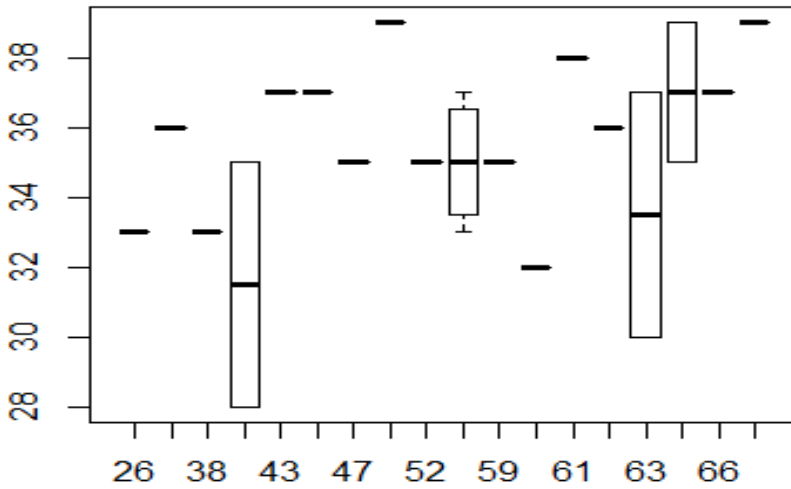
## Appendix IV: The lexical items in the word-recall task

<u>Word</u>	<u>Type</u>	<u>Occurency (Yes/No)</u>	<u>Frequency (COCA)</u>
buddy	noun	Yes	high
paradise	noun	No	med
quirky	adjective	Yes	low
costume	noun	Yes	med
candle	noun	no	low
lover	noun	Yes	high
gross	adjective	Yes	high
victim	noun	Yes	high
guitar	noun	No	high
flask	noun	No	low
killed	verb	Yes	high
explore	verb	no	high
abyss	noun	No	low
surfing	verb	Yes	low
backpack	noun	No	low
sassy	adjective	Yes	low
genital	adjective	Yes	low
repetition	noun	Yes	low
crew	noun	Yes	high
disgusting	adjective	No	low
record	noun	Yes	high
miracle	noun	Yes	med
shoved	verb	Yes	med
sheep	noun	No	med
awards	noun	Yes	high
terrific	adjective	No	med
retaliate	verb	No	low
bury	verb	No	low
thorn	noun	Yes	low
solemn	adjective	No	low
ponder	verb	No	low
eager	adjective	No	high
decoy	noun	Yes	low
messiah	noun	Yes	low
fatherly	adjective	No	low
incognito	adjective	Yes	low
company	noun	No	high
arrested	Verb	Yes	high

miracle	noun	Yes	med
puked	verb	Yes	low
hebrew	adjective	Yes	low
error	noun	No	high
paycheck	noun	Yes	low
argue	verb	No	high
invasion	noun	Yes	high
considered	Verb	No	high
unique	adjective	No	high
uranium	noun	Yes	low
ghastly	adjective	Yes	low
betray	Verb	Yes	low
poop	noun	Yes	low
pilot	noun	Yes	high
munch	verb	Yes	low
lawyer	noun	No	high
boiler	noun	No	low
surgeon	noun	No	med
paraplegic	adjective	Yes	low
compete	verb	No	high
bigfoot	noun	Yes	low
diner	noun	Yes	low

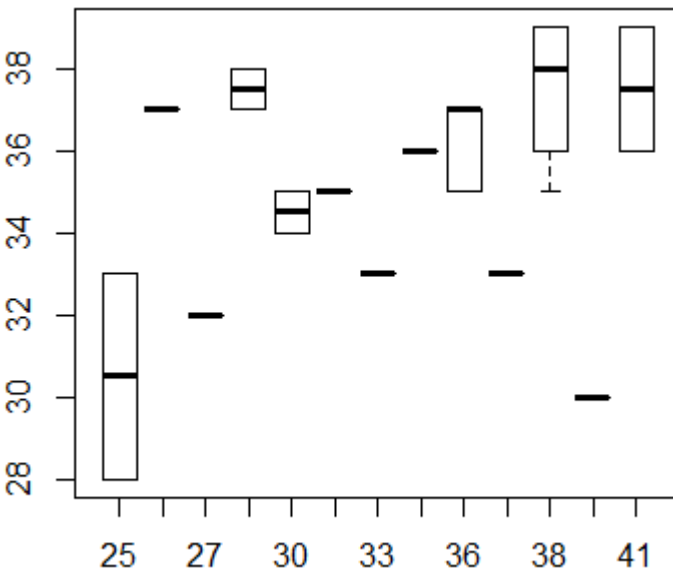
## Appendix V: Boxplots of linear mixed models

Figure 1. | Boxplot of the relationship *Word-definition task ~ vocabulary*



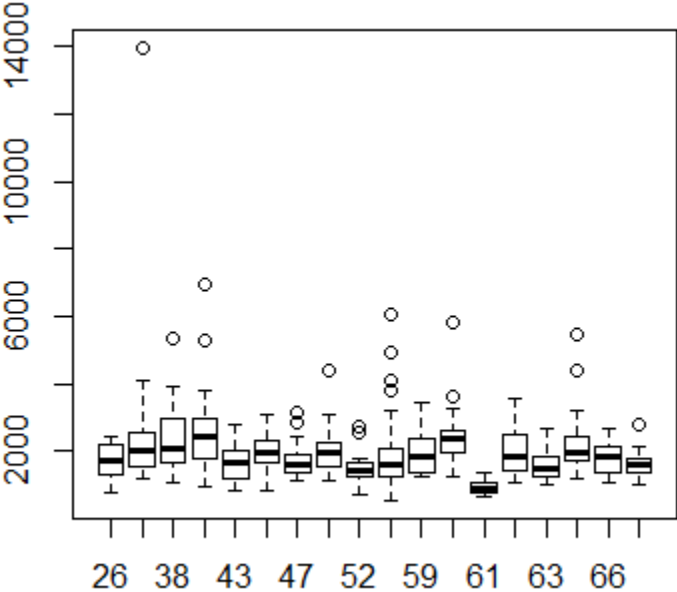
Note: X-axis = result on vocabulary task, y-axis = result on word-definition task.

Figure 2. | Boxplot of the relationship *Word-definition task ~ grammar*



Note: X-axis = result on grammar task, y-axis = result on word-definition task.

Figure 3. | Boxplot of the relationship *Reaction time ~ vocabulary*



Note: x-axis = results on the vocabulary test, y-axis = reaction time measured in milliseconds

