

Henrik Eye

***The Impact of Subtitles in Second Language
Acquisition***

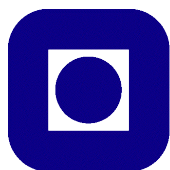
**An experimental study on Norwegian
upper secondary students**

Master's thesis in English

Trondheim, 2016

Supervisors: Mila Dimitrova Vulchanova and Giosué Baggio

**Department of Modern Foreign Languages
NTNU**



**NTNU
Norwegian University of
Science and Technology**

Abstract

In this study, we sought to find and measure the long-term effects of subtitled second language audiovisual material on second language acquisition, and how significant the potential effects were. The participants of this study were students from a VG3 level in upper secondary school class who were divided into three groups. Two of these groups were experimental groups who watched four episodes of Family Guy with either English or Norwegian subtitles, while the third group, being a control group, watched the material without any subtitles. This master's project has been a continuation of Ingrid Elisabeth Nufsfjord Kvitnes and Lisa Marie Grønn Aurstad's theses on the short term effects of subtitles in second language acquisition. This study looks more extensively on the long-term effects in the same area.

We found that viewing intralingual subtitles, in this case English, benefitted students on word recall aptitude, and that vocabulary and grammar proficiency had an effect on word definition skill. Furthermore, we established that word frequency was a predictor for accuracy on the lexical recall task, showing that higher frequent words are more easily recalled after a longer period of time after being exposed to the material (circa 1 month) than lower frequent words. An additional finding was that both proficiency with vocabulary and grammar were linked to word definition ability, meaning higher expertise with vocabulary and grammar establish better scores on word definition tasks.

Acknowledgments

I am thankful for being able to write about the subject of subtitling, which I would have never been able to do without the help of the Department of Modern Foreign Languages. I am especially grateful to my supervisors, Professor Mila Dimitrova Vulchanova and Associate Professor Giosué Baggio for advising, supervising and teaching me helpful skills in order to conduct experimental research of this magnitude. I would also like to thank my partner in this project, Erlend York, who has been of great help planning and conducting the experiments. I also thank Emily Fischbeck Luthentun for proofreading and supporting the project through this year. This study is a continuation of the research conducted by Ingrid Elisabeth Nufsfjord Kvitnes and Lisa Marie Grønn Aurstad's theses (Vulchanova et al.), compelling me to thank these two for having paved the way for a thread of discussion and further exploration in this field of research. I would also like to thank my contact teacher from the class used in this study for helping me plan and conduct the testing. Lastly, I want to thank Sobh Chaboun for building the lexical recall task used to test the participants of this study.

Henrik Eye,

Trondheim, May 2016.

Table of contents

Preface	9
1. Introduction	11
2. Theoretical background	11
2.1. Second language learning	13
2.2. Grammar and vocabulary	15
2.3. Multimedia in second language acquisition	16
2.4. Subtitling and dubbing	17
2.5. The role of input and interaction in SLA	21
3. Method	25
3.1. Aims of the present study	25
3.2. Participants	26
3.3. Procedure	28
3.4. Analysis	31
4. Results	33
4.1. Vocabulary and grammar tests	33
4.2. Word definition task	35
4.3 Lexical recall task	39
5. Discussion	45
5.1. Vocabulary and grammar	45
5.2. Word definition task	45
4.3 Lexical recall task	47
4.4 Limitations of the study	50
5. Conclusion	51
Appendices	58

Preface

The subject of audiovisual subtitling was selected as a topic for this thesis in the context of me being a teacher of English. Language is a particular interest of mine, and multimedia material's potential for facilitative effects on language acquisition was a field I wanted to learn more about, principally due to its countless factors still open to scrutiny. This thesis, with its research and experiments will help me as a language teacher understand more about the frameworks of language acquisition, and thus let me explore some of the vast techniques beneficial in learning a second language.

This study has been an attempt to test and discuss possible significant long-term effects of using subtitles, both in terms of interlingual translation of the video material's message and an audio-to-text transcription (intralingual translation) of the language.

The Norwegian Directorate for Education and Training is responsible for development of kindergarten, primary and secondary education, with the latter being my field of expertise. This study is highly relevant for my career as a teacher, in that it explores a multitude of second language aspects within the field of linguistics. In fact, one of the competence aims of the course 'Social Studies English' mandates that students should "analyze linguistic tools in texts in dissimilar genres and assess their impact" (UDIR, 2006). Subtitling is indeed a linguistic tool, and working with the impacts of it mirrors the aims of this thesis. Students are encouraged to work with, and to be critical of these tools in order to maximize their learning outcome. The latter is one of the aims of studying language acquisition.

1.0 Introduction

With schools rapidly becoming more technologically dependent and digital skills being required both as a form of competence for students, but also for teachers, it becomes necessary to embrace such a shift in order to fully take advantage of new methods of learning. As a teacher of English, I have had many good experiences working with video material in the classroom, but often when regarding usage longer audiovisual clips, the question of subtitles will usually appear in some form. Some students feel that the subtitles distract them from the plot, and if it is in a different language than the audio, it becomes even more distracting. Others feel a more dependence on subtitles, with reasons being vast and different; usually it is to better clarify what is said when lines appear to be spoken in a rather challenging dialect, tone or volume. In short, the use of subtitles is for comprehension of the message. In any case, using subtitles becomes a sort of a dilemma if one does not know whether using it may turn out to have a beneficial, or in contrast, a harmful effect on the student's language learning capabilities.

Isabel Borrás writes,

“Opponents of the use of subtitled video in foreign/second language teaching argue that the presence of subtitles is distracting and that they slow down the development of learner's listening abilities. Proponents of subtitles, on the other hand, content that subtitles may help develop language proficiency by enabling learners to be conscious of language that they might not otherwise understand” (Borrás, 1994, p. 61).

The concern is therefore the nature of the subtitles. One could argue that using English subtitles with English video material could enhance the language learning because the viewer both sees and hears the language, being exposed to multiple instance of the same language, thus enabling them to pick up pieces of information otherwise missed by those not using subtitles at all. In the same way, one could say that using subtitles that differ from the source language (intralingual subtitles) could create a dissonance between these languages and potentially result with the viewer either not learning anything, or possibly even interfering with language learning.

To fully understand how subtitling can cause an effect on language acquisition, one must observe studies where this has been the matter at hand. Mitterer and McQueen (2009) conducted an experiment on native Dutch speakers, measuring differences in various second language proficiencies after exposing participants to subtitled audiovisual material. In this

study, Dutch students learning English implicated negative effects when exposed to English video material where the audio was in English and subtitles in Dutch. In contrast, there were positive effects when using English video with English subtitles (Mitterer, 2009). Mitterer writes that the Dutch students were instructed to view videos containing unfamiliar regionally-accented English, with or without subtitles. This master's thesis will look at English and Norwegian subtitles. Mitterer's test was therefore conducted in order to tell whether subtitles help or hinder adaption to an unfamiliar regional accent in a second language (ibid). This thesis is in many ways based on a testing similar to that of Mitterer's, but will for most parts concern plot-dependent variables of language as we are interested to learn more about what long-term effects of exposure to subtitles have on the second language learner.

The research question that I am constructing is therefore: *What are the long term effects of exposure to subtitles in audiovisual material for second language acquisition?* Additionally, a secondary aim of the project is to investigate any differences in these long-term effects in different levels of proficiency.

The hypothesis to this research question is that we expect to find significant long term effects of using subtitled audiovisual material.

Kvitnes and Aurstad (Vulchanova et al.) did not find any significant long-term effects in their study, but it is important to note that their study limited itself to one single episode of *Family Guy*. I want to show one episode each week for a total of four weeks, making sure that the students are exposed to enough audiovisual content over the course of an approximate month. It therefore becomes logical to project a different outcome due to the change in amount of stimuli and exposure time.

2.0 Theoretical background

2.1 Second Language Learning

The object of Second Language Acquisition (SLA) research is to study individuals and groups who are in the process of learning a language subsequent to learning their first one as adolescents (Saville-Troike, 2006, p. 2). It is the native language (hereby L1) that is most prominent with the speaker, meaning the language the speaker is most proficient with. In this thesis and the experiments tested on the upper secondary students (VG3, specifically), the L1 in question will be Norwegian, while the second language (L2) will be English.

Trying to understand the process of SLA, we seek to answer three basic questions: (1) ‘What exactly does the L2 learner come to know?’ (2) ‘How does the learner acquire this knowledge?’ and lastly (3) ‘Why are some learners more successful than others?’ (Ibid). There are no simple answers to these questions, which suggest how complex and intricate the nature of SLA is. There are many factors that go into play when learning a language, which are more easily studied individually. Multimedia’s effect on the SLA is therefore one example of an instance that can have either positive or negative effects on the language learning, depending on different variables, both in terms of inputs and outputs. In any case, it might be difficult to try to answer the three questions with only one case study, as language learning is far more complicated for those questions to be solved with simple answers. This individual study will focus on one single form of impact on the SLA, namely audio-visual material. This study will therefore not be able to answer these questions by itself, but will be helpful in expanding the discussion of SLA. The impact this study could have is to provide evidence that suggest certain effects and impacts audio-visual material have on language acquisition.

Vivian Cook describes SLA research as something which draws aspects of linguistics, psychology, sociology and education into one single field (Cook, 2008, p. 6). SLA researchers are often exclusively interested in one single instance of language learning, whether it being phonology, vocabulary, grammar, social impacts or other elements than can be analyzed and discussed in the scope of language learning. Foster-Cohen (1999) notes that L1-L2 connections have not always been explored, leaving big questions unanswered in two fields that are closely related (p. 4). In fact, studies on audiovisual subtitling and its language

impacts is a relatively young branch in the scope of language studies. Foster-Cohen points out the lack of child second language studies (p.7), although now in 2016 considerably more covered, and argues that such a study can offer broader knowledge on the concept of a ‘language instinct’ in humans, an aspect on language processes that is continuously debated. Nevertheless, language acquisition is a scientific field with many uncharted areas which calls for extensive examination.

While there are numerous factors that decide how an individual learns a language, one might still find certain patterns with speakers that connects them in a logical way. Running tests to find specific correlations, like testing for proficiency when exposed to a certain language-embedded element, can therefore be helpful in order to investigate the nature of their proficiency; what influences their proficiency and what consequences they incorporate. Cook argues that all successful teaching depends on learning; that there is no point in providing entertaining, lively, well-constructed language lessons if students do not learn from them (Cook, 2008. p. 6). In the case of using multimedia material in the English classroom, the teacher should therefore be clear on his or her aspirations for what sort of learning outcome they want the audio-visual materials to project. While multimedia entertainment can provide positive results with the language learner, there needs to be a sense of logical function or intent behind the use of such elements. If there is not any purpose behind viewing audiovisual material, any positive impacts on learning will naturally be arbitrary as a result. The students could have been instructed beforehand to pay special attention to various scenarios and nuances, for example. However, a consequence of using audiovisual material with a specific and clear learning goal is that the material becomes didactical, and thus loses the spark gained by not being conventional “boring” learning material. This will be discussed further in chapter 2.3.

The origin of SLA as a scientific field originated embedded with behavioristic tradition (Johnson, 2004). It is also closely related with contrastive analysis (CA), which in many ways influenced not only SLA theory, but also L2 classroom teaching. In Europe, CA was initially viewed as an integral part of general linguistic theory, and the idea was to understand and explain the nature of natural languages. In the US, CA had strong pedagogical roots, adhering to the scientific tradition of its time – namely behaviorism (ibid). To fully understand the concept of SLA research and theory, one has to take its origins into perspective. Knowing that second language acquisition theory stems alongside with the theoretical paradigms of behaviorism, one can more easily comprehend its historical basis and development alongside

with shifts in pedagogical theories, as there are apparent correlations between them. In any event, when studying the roots of SLA research, one must take into consideration the development of these practices, alongside the practice of language teaching and learning, as mentioned by Plass in chapter 2.3. The bottom line is that SLA research and theory has undergone several important changes in the past 60-70 years, which has helped shape the practice of studying the complexity of language acquisition.

2.2. Vocabulary and grammar

Zimmermann (1997) calls vocabulary “[...] central to languages and of critical importance to the typical language learner (p. 5). Throughout the nineteenth century, vocabulary was taught through etymology and definitions, much due to the connection between them (ibid). Later on, the grammar translation method, involving preparation of students to read and write classical materials and to pass standardized exams was globally predominant. Historically, grammar and vocabulary teaching have been the center of second language learning, because they are the most principle elements of a language; one considering form and the other content. Among different types of linguistic knowledge that constitutes successful reading comprehension, grammar and vocabulary have received considerable attention in linguistic research. (Zhang, 2012, p. 558) The two, as Zhang correctly argues, are fundamental in building local, as well as global coherence, and thus affecting comprehension. If we accept that grammar and vocabulary are foundations of language comprehension, it becomes crucial to measure these two factors to illustrate to some degree the language proficiency within subjects. One does this in order to view participants’ level of language proficiency and treat data output generated by subsequent testing on the same subjects. This is what Zhang calls an ‘index of L2 proficiency’ (ibid, p. 259). This is a background of an estimation of language aptitude, not to be confused with in-depth analysis and assessment of language proficiency. It is generalizing because as stated in chapter 2.1, language is very complex, and measuring linguistic handiness is virtually impossible due to the seemingly unlimited factors that constitute psycholinguistics.

In the context of being a teacher of language, López-Burton (2014) writes,

“Many teachers feel that teaching grammar or vocabulary is their main role. We can’t teach fluency – that is a process that just happens – but we can teach *something*. We can teach rules!” (López-Burton, 2014, p. 24)

Appropriately, rules are what make up the two. Without rules in a language, confusion develops, haltering or fully disabling the possibility for communication. This allows us to perceive grammar and vocabulary as something steady and concise – something measurable. Testing for vocabulary and grammar proficiency enables the researcher to get a grasp, although relatively limited, on a language learner’s aptitude with the particular language. In the context of this thesis, grammar and vocabulary tests will be constructed and conducted in order to measure second language proficiency as a manner of background of the participants. If the scores differ in any way between respective groups, any subsequent tests would naturally be influenced by their skill with the second language.

2.3 Multimedia in Second Language Acquisition

Using a language, either the first or second, has as its goal to communicate ideas, maintaining social relations, and creating discourse, all which requires several different core competencies (Plass, 2005, p. 467). Some of these competencies include reading, speaking, writing and listening, all important factors in the process of learning a language. The use of multimedia, notably audiovisual material can fulfill some, but not all, of these criteria. While one could argue for audiovisual material’s potential in the English-speaking classroom, one must not overestimate its value as a basis for teaching, solely relying on such material to cover all fields language teaching encompasses. Nevertheless it is correct to state that multimedia’s role, both in general society, but also in the premise of the school, has become more prominent in the past decades.

Plass writes:

“However, the approaches taken to the teaching and learning of a second language have changed considerably over the past 40 years, loosely mirroring the development of psychological theories and models of teaching and learning” (ibid).

The development of technology, psychological theories and teaching formats alike, has changed the way teaching and learning is viewed and practiced, with new teaching plans emerging to comply both with these shifts and the ideals of politicians responsible for

creating these plans. One of the most prominent of the newer changes in the past decade is the digital competence aim, which is defined by Udir as a skill that involves using digital tools, media and resources for the purpose of solving practical tasks, gathering and process information and to create digital products and communicate (Udir, 2012, web). One of the general principles of Norwegian schools has as its aim to educate individuals to prepare them for society, such a skill is undeniably important to thrive in a modern, western society where such skills have many practical uses, both professionally and recreationally.

Iva Baltova writes,

“Different kinds of input (oral speech, written text, and visual cues such as setting, objects, people, actions etc.) have unique qualities of their own that may facilitate the comprehension and learning of L2 vocabulary and content.” (Baltova, 1999, p. 1)

These elements are all available within the same medium, and thus collaborate to affect the viewers' language proficiency, although usually arbitrarily due to the non-didactical nature of television shows. It becomes therefore especially important to keep other probable factors in mind when measuring the case of subtitles, as it is usually a collectiveness of factors that altogether have an impact of the learner's proficiency. Whether subtitled audiovisual material has an effect on vocabulary or other instances of language aptitude, one must consider that subtitles alone cannot be isolated as a sole operator in the premise of language learning through viewing multimedia material. This study aims to measure whatever effects subtitles have on language acquisition, and as in the study of Vulchanova et al., a control group which will view episodes without subtitles will be important in order to measure whether the presence of subtitles for the two other groups, either intralingual or interlingual, create a different output.

2.4. Translated subtitling and dubbing

Audiovisual translation is not one type of translation, but rather a set of different strategies in which the translator can move content in audiovisual material by shifting elements in the various channels of the medium. Effectively, there are seven types of audiovisual translations: interlingual subtitling (text on the screen, translated from the audio channel of the original language to another), bilingual translation (in countries like Belgium where subtitles are provided in two languages), intralingual subtitles (transcription of the audio into

the same language), dubbing (where the voice track is altered, usually translated into a different language for comprehension), voice over (mainly used for documentaries, commercials and interviews), surtitling (subtitles which can be projected, for example, above stage of an opera), and finally audio description (intralingual commentary on the action for the visually impaired) (Munday, 2012, p. 271). In this study, the intralingual subtitles used in the experiment will be in English, while the interlingual subtitles will be in Norwegian. The control group will not view episodes with any type of audiovisual translation.

Many critics of the use of translated subtitles in audio-visual material tend to attack the same aspect, namely the fact that it has undergone translation from its original language. Some even go as far as calling such subtitling ‘abusive’. To some, the original, foreign, object – its sights and sounds – is available to all, but it is easily obscured by the graphic text through which one necessarily approach it (Nornes, 1999, p. 18). Knowing that there exists critique on subtitling before conducting testing on its effects on second language acquisition is helpful in the process of viewing general approaches to subtitling, both from the perspectives of researchers and learners alike. Additionally, reading these kinds of opinion in advance of researching the impacts of translated subtitling gives an impression that there are certain concerns of negative effects regarding its use. If one takes the studies of Kvitnes and Aurstad into this context, one will agree that their findings reported a problematic result for those proposing positive effects of translated subtitled audiovisual material. In light of this, it is only natural for me to predict relative negative impacts on the L2 learning regarding the students exposed to L1 subtitles with the L2 video material.

Another issue with subtitles is what Gottlieb (1994, referenced by Munday, 2012, p. 278) describes as an “overt translation”: the visibility of the title is a built-in part of the activity of viewing the rest of the material. In other words, it becomes a secondary task for the viewer to process even more information if he or she is expected to pay simultaneous attention to both elements. Diaz Cintas and Remael (2007, p. 40, referenced by Munday, 2012, p. 279) use the term ‘vulnerable translation’ about subtitling in general and argues the following:

“Not only must subtitles respect space and time constraints, they must also stand up to the scrutiny of an audience that may have some knowledge of the original language” (ibid).

The audience, with some degree of knowledge of the source language, has an expectation to the subtitles, which can have various consequences if some of the material is omitted or reduced, which again might cast doubt on the quality of the subtitles. The vulnerability is not

as represented in other types of translations, because of the lack of comparison through the simultaneous channels of text, audio and video, meaning that there is more pressure put on the agent of subtitling (ibid). This pressure could as a consequence provide better subtitles, because of the sheer amount of expectations from the audience's side. Dubbing, which is my next point in this discussion, is something that in appearance seems far less "overt."

A common standard alternative to subtitling is the practice of dubbing, which incorporates auditory translation with the usually unaltered video channel. Dubbing is a fairly common practice in countries such as France, Spain, Italy, but also increasingly in Norway, as opposed to subtitling. The practice of dubbing is relatively more expensive than subtitling, due to the need of voice actors and more work required to finish up the translated product. However, dubbing effectively manufactures a product more relatable to the L1 audience, usually children in Norway. In a broader sense, it is the largest and most globalized linguistic communities that choose to dub instead of subtitle movies and television shows, while smaller linguistic communities prefer subtitling (Gottlieb 1996, Luyken 1991, as referenced in Blystad and Maasoe, 2004, p. 6). Despite dubbing being a more expensive industry than subtitling, the ever growing Norwegian film industry in cooperation with the global film industry (namely Hollywood) has chosen to rely more on dubbing than on subtitling, since it generates larger revenue for the industry. Dubbing, in contrast to subtitling, has no beneficial value in the scope of second language learning, which incidentally is why it is important to bring into the discussion. Subtitling may aid in facilitating second language acquisition, thus making it the better alternative.

Sheila Turek (2010) states that standard conventional subtitling in audiovisual material usually confines within 40 characters per line, including spaces with a minimum display time of 1.5 seconds and maximum display time of four to five seconds (p. 560). This makes subtitles rather condense in appearance, and some utterances might be left out. This can include repetitive and thus redundant words that would otherwise hijack the positions of more important lexical items in the limited frames of on-screen subtitles. Another challenge with translated subtitles is that some cultural instances might be altered rather extensively. Turek writes,

"Additionally, filtering messages through another society's idiom can render various nuances of the original inaccessible on many levels to viewers reliant on subtitles for meaning. Several

researchers in the burgeoning field of audiovisual translation studies have documented the difficulties of cultural transmission through subtitles [...]” (Turek, 2010, p. 550).

In the event of using *Family Guy*, it becomes all the more important to be aware of the potential cultural differences that becomes lost in translation subject to a degree of difficulty when translating. The plot of the series takes place in Rhode Island, United States of America, and the main cast is portrayed as a stereotypical American family (i.e. obesity and illiteracy issues, and generally characterized as asinine). Many of the jokes and gags, in many situations conveyed through an anecdotal scene, as is custom for the series, are culturally significant for an American audience. One could argue that Norwegian culture has adopted much from America, thus providing sufficient cultural context for such a show to be viewed and understood by a native Norwegian individual. However, the plot is occasionally tied-in with American politics or American celebrities, rendering various antics and plot-lines inaccessible to a foreign audience, if deemed to be too culturally padlocked. It is consequently essential to be aware of these cultural differences, as they might disrupt a quantitative study if not approached accordingly. With this in mind, words selected for the lexical recall task will not incorporate culturally restricted words. However, the word definition task includes some rather American phrases, such as “a red state” (meaning a state where the popular vote usually goes to the Republican Party). Such instances will be allowed in the word definition task because of the abundance of culturally loaded plot-lines in *Family Guy*. After viewing the episodes, we aim to find out whether the participants are familiar with these terms and phrases. The lexical recall task requires words to be simple for the sheer activity of recalling the words; more complex words can distort the data output (i.e. too many syllables).

An unintended advantage of subtitling, especially in contrast to the practice of dubbing, is that learning effects may occur (Koolstra, 1999, p. 52). A panel study conducted by Cees M. Koolstra has shown that television’s impact on children’s reading skills may improve children’s development of decoding. Reading subtitles provides extensive practice in decoding words (ibid). Learning through subtitled audiovisual material in the comfort of one’s own home is largely an unintentional learning experience. One could argue that it is a positive phenomenon since the process of learning then becomes free of didactical intention, meaning the individual acquires knowledge constructively on its own. In a classroom, the student is usually aware of the didactical processes, knowing that the message conveyed by the teacher is uttered in relation with the goal of a learning outcome. This is especially

relevant in regards to unmotivated students who tend to distance themselves from learning experiences at school, because they feel they are forced to learn something rather than wanting to learn something.

An early study of 32 English speakers learning German who had watched German video material with English subtitles admitted to having “given up” on understanding the foreign language just a few minutes into viewing German film material and instead relied solely on English subtitles for comprehension (Froelich, 1988, p.199). While this example is different in the way that the English language for Norwegian students is fairly easily comprehended through speech alone, German for English speakers tends to be much more of a challenge. Norwegian students encounter more English every day than Native English speakers typically do with German, unless confining themselves to a German-speaking living situation. The issue presented by Froelich nevertheless illustrates the challenge of interlingual subtitling as a basis for language teaching, problematizing the area of dependency of subtitling with the viewers. While there is an acknowledged potential for language learning with either interlingual or intralingual subtitles, it is the reliance of subtitling, thus ignorance of the spoken language in the audiovisual material, that becomes the problem in the context of language learning. As long as the teacher is aware of this matter, he or she may appropriately present audiovisual material in a way that streamlines language acquisition, relative to social and linguistic contexts.

2.5 The role of input and interaction in SLA

In a broad perspective, input can be viewed as an interaction. In a more technical classification and predominantly in relation to SLA, Marjolijn Verspoor characterizes input as processes that involve strategies and mechanisms that help make connections between particular language forms and their meaning during comprehension (Verspoor, et al., 2007, p. 1). In language learning, input is the exposure to degrees of language, either written, spoken or in visual forms. The role of audiovisual material, with the inclusion of subtitles will therefore cover all three aspects. It does this by visually presents its content as animated feature on the screen, with spoken dialogue from the audio channel, and lastly subtitles as written text that appear with the animated pictures on the screen. These three elements are therefore three individual sets of inputs, which all together convey a correlated message to an

audience. If one is to study the input of subtitles as isolated as possible, one would have to detach the particular input from the two other, which only reduces it to pure text, thus eliminating it as a definite subtitle. In short, subtitles are only defined as such if it is in accompaniment with its other two other co-operating factors. This becomes quite a conundrum in an experimental research design, as one cannot possibly study subtitles in isolation. When studying subtitles, one is at the same time studying the rest of the audiovisual material, meaning that this study has to take the other two elements into consideration when building, planning and conducting experiments to measure for language impacts. Such factors will be discussed further in the method chapter (chapter 3).

Lev Vygotsky and his colleagues were one of the earliest pedagogy theoreticians who worked with interactionist research. The role of language in learning was one of Vygotsky's key ideas in learning processes in general aspects of learning, much because of how knowledge is exchanged in interactions between human individuals. Susan M Gass writes:

“Extensive empirical studies of input and interaction explored the ways in which learners manipulated their interlanguage resources when asked to make their messages more comprehensible.” (Gass, 1998, p. 299).

Gass points out that the manipulations led learners to restructure their interlanguage toward greater accuracy and complexity (*ibid*). The role of input has therefore become critical for SLA research, but is by all means a product of research in itself that is constantly in development, as Gass notes. The sheer wealth of information produced in this line of research has since the 1980s challenged SLA researchers in terms of development and internalization. Despite promising results of the early research on input in SLA, the effect of interaction on acquisition has remained a complex issue (*ibid*). It is not surprising, considering the sheer abundance of psycholinguistic factors that are ever so difficult to pinpoint and map to create a standard for any type of psycholinguistic measurement. The brain remains a complex issue, something researchers must take into consideration when conducting SLA experiments and research.

Fortunately, despite the complexity of psycholinguistic study, there are certain patterns in the human brain when it comes to language that can be pinpointed and isolated to some extent. For example, certain word classes, like nouns and verbs (content words) tend to be processed faster in the brain than other, more complex lexical items. This is due to the fact that these types of words are easier for the individual to relate to something concrete, making human

subjects have faster response time when encountering verbs and nouns. This is a well-known matter within psycholinguistic research, labeled as the Primacy of Meaning Principle, meaning that learners process input for meaning before they process it for form (Lee, 2009. p. 4). The word classes selected for the lexical recall task in my study will consist of nouns, verbs and adjectives. This is not solely to adhere to the Primacy of Meaning Principle, but rather because they are content words that describe something specifically and the test was created in order to measure whether the subject could recall the uttered lexical item in the episodes. This would be completely impossible with form words that bear no resemblance to specific episodes. How could one measure whether a subject had encountered a specific word in a distinct situation if the chances that he or she would have come across the word any other place? There is no guarantee that a lexical recall task will be free of interference, but such issues are all the more important to be aware of when constructing and conducting one. There is therefore a priority to select English words that a native Norwegian would be less likely to encounter in other situations, including in the English speaking classroom. It needs to be addressed in order to produce cleaner data. The words selected for the lexical recall task are also prominent in the episodes due to the fact that they are often linked with jokes and gags, being scenarios that are expected to be more easily remembered. Other instances are otherwise relevant to the episodes' plots, making them more noticeable than words that might be regarded as part of everyday discourse. Family Guy is a situation comedy cartoon, making most dialogue either related to the plot because of the limited amount of screen time (usually around 22 minutes), or some type of humorous spiel in form of jokes and other anecdotal sketches. This is yet another argument for why Family Guy was chosen, say in contrast to other television programs, because a lexical recall task can be created and conducted with less of a risk that the words uttered might be missed by the audience.

3.0 Method

Throughout the process of planning, developing and conducting the experiment, as well as collecting and reading data, I have collaborated with another master student who is writing a thesis related to this one. The major distinction between the two projects is the age groups involved; the participants in my study were upper secondary VG3 students, while the other project used students from VG1. The intention behind testing two different age groups was to measure similar effects but with an expectation of different levels in proficiency. We both shared the same supervisors in this project and have collaborated and aided each other with both technical and practical matters. Kvitnes and Aurstad's theses (Vulchanova et al.) shared this correlation as well, which is why it is important for us to keep this noted, and to distinguish the use of "I" and "we" when referring to either myself or both master students.

3.1 Aims of the present study

The purpose of this study is to look for long-term effects of viewing L2 audiovisual material. To find out whether viewing multimedia clips with Norwegian subtitles, English subtitles, and no subtitles have any consequences for the L2 proficiency, a deductive method was necessary in order to truly respond to the research question. Additionally, an experimental approach was chosen because such a study with this specific question at hand has not been conducted in the context of research on Norwegian upper secondary students. While this study in many ways replicates the thesis designs of Kvitnes and Aurstad (Vulchanova et al.), the major distinction is this study's scope on long-term effects, which allows me as researcher to inspect new elements in SLA research. Because this subject of this latitude is relatively new in the field of psycholinguistics, it becomes important to tread lightly with the results produced from this study, and to scrutinize the data output in depth. With relevant theory to support and help shed light on the abundance of language-specific nuances within this study's datasets, I will analyze and discuss the findings appropriately.

Furthermore, the role of digital testing is important to this study, namely an E-prime built program used to conduct a lexical recall task. A significant advantage of building a digital test like this is the possibility to test for reaction time within and between subjects. By measuring reaction times, we can see whether independent variables such as word frequency, word occurrence and word class have an effect on time spent responding to each query. The

raw results will be restructured in Excel and analyzed in R to scrutinize effects and significances. All results relevant to this study will be reported in chapter 4.0.

3.2 Participants

The participants of this study consisted of 21 students, all from one VG3 class. Most of the students were 18 year olds (age range 18-19) and the testing was done between December 2015 and February 2016. 13 of these participants were male, and 8 were female. One should note that these students are in the final year of their study program (studiespecialisering in Norwegian), with English being an optional specialization course. This becomes especially important if one is to compare the results of the other master student's testing, as the level of proficiency in the higher level students is expected to be substantially higher than that of the lower level VG1 students. English is an obligatory class for the VG1 students, and one should take this into consideration, as well as their age, when analyzing the data produced from both of these studies.

In hindsight, one single class consisting of 21 participants may not prove to be as representative as Ingrid Kvitnes' total of 65 participants. Initially, 24 students agreed to participate in the study, but two of these students did not participate in enough video viewing sessions. Another student responded to have various language learning difficulties, forcing us to remove them from the study. The outcome of this is that the three divided groups still had an even number of participants between them. However, because of the moderately low number of students per group, there is naturally a higher chance of indiscretion with the data from the results of the testing. One needs to take this in to account when viewing and interpreting the product of the tests' outcome. One might argue that the scientific method of this study becomes, in a sense, more qualitative because of the fewer numbers of participants, and with the inclusion of a digital test that looks at many different variables and then produces a large abundance of data accordingly. Granted, this study is by definition of a quantitative nature due to analysis of statistical data, but it is nonetheless important to add into the equation.

The participants' teacher was my contact person in the process of planning and conducting the experiment. I gave the teacher a list that divided the students into three groups. I chose to call the participants by numbers (000-024), in which the teacher would assign one to every

individual student. By doing this, I would not know the students' names in combination with their codes; only the teacher had access to that information. Unfortunately, there was one student with language learning difficulties, thus unable to participate in the study. Two other students had not been exposed to enough stimuli (absent from viewing sessions), forcing me to remove them from the study. The total amount of participants was already substantially lower than those of Kvitnes and Aurstad's student participants, so it became a worry that some of the participants had to be excluded for various reasons. Additionally, because each group consisted of maximum 7 people, I had to make sure every test, either written or digital was done thoroughly and with satisfactory supervision. The other master student helped me supervise and conduct the digital testing, while I helped him with his. In that way, we could make sure everything went according to plan.

Group 1 – Norwegian subtitles

This group consisted of 7 students, originally 8 before exclusion of a student who did not participate in enough viewing sessions. All members of this group were Norwegian native speakers, but one responded to be bilingual (with Norwegian as mother tongue). This group watched the four family guy episodes with Norwegian subtitles.

Group 2 – English subtitles

This group consisted of 7 students after the exclusion of a student that did not attend any of the tests in February 2016. All of the students responded to be native Norwegian speakers. The group watched the Family Guy episodes with English subtitles.

Group 3 – Control group

This group also consisted of 7 students out of the original 8 after having excluded one participant with learning difficulties. All of the participants responded to be native Norwegian speakers. This group watched the episodes with no subtitles.

3.4 Procedure

Background information questionnaire

The background information sheet (appendix) was handed out subsequently after the participants had been officially informed about the project and signed the consent forms to be a part of it. The contents of the background information questionnaire were elaborated orally in plenum and given out to each individual student. The form contained closed and open-ended questions about the participant's language background and any other information related to the study. One of the more important questions, apart from native language, were concerning the amount of time spent being exposed to second language video material or video games, which would be potential factors for their L2 proficiency. Some of the questions required the participant to respond with either "yes" or "no". Others required them to choose an alternative ranging from "never" to "every day". In the analysis of the data collected, the answer alternatives were converted to numbers instead. For instance, when asked to rate their own proficiency with the English language, any instance of "fluent" would be converted to a 4, whereas "basic" would correspond with a 1. The questionnaire also asked for other factors of language acquisition, and finally any potential mentally or physically handicaps that would have an impact on their ability to speak, write or comprehend language. If there were any such diagnoses and deficiencies, the participant would be excluded from the project.

Vocabulary and Grammar tests

Before the students were allowed to view the episodes, they agreed to be tested with a grammar- and a vocabulary test, both digitally and online through a web browser. This was conducted in order to measure the average level of grammatical and vocabulary proficiency. The students had brought their own personal school laptops, but were supervised by both their teacher and I as they took the tests. The students were disallowed from communicating with one another, as the same principals as taking a graded test applied. They were told that none of the tests conducted would affect their grades in any way, and the teacher would never see their scores. This would insure them that I would only use the data for scientific purposes, and also making them more relaxed about the situation. When finished with the test, the participants were instructed to not touch their computers, and rather let me come and oversee

their results, which were immediately written down at the end of their background information sheet, all under supervision. The grammar test consisted of 50 sentences where the students would choose the most suitable expression to fill in the blank of every sentence. There were 4 options for every task, and each correct response awarded one point, giving a score from 0 to 50. The vocabulary test required students to press two individual buttons on their computer's keyboards, one for yes, and one for no, to respond to whether they had heard of various singular words. The online program would show one word at the time on the screen and wait until the participant had input either response key to proceed to the next word. The result given was a percentage, therefore 0 to 100 %. A small amount of the appearing words were nonsensical, non-existent control words, which would automatically subtract the score of the participant if "yes" was input. Accompanying the percentage score was also a sentence that would inform the participant that the score was of an X level for a beginner, intermediate or a native speaker of English, where X ranged from low to high. The results of the vocabulary tests, just like the grammar tests, were overseen individually and written down correctly at the end of the background information sheet.

The vocabulary test can be accessed at <http://vocabulary.ugent.be/> and the grammar test at http://www.examenglish.com/cpe/CPE_grammar.htm.

Word definition task and lexical recall task

Approximately 2 months after the initial viewing of the four episodes, I returned to the school to conduct a final set of testing. The first test was a word definition test (appendix C), which consisted of a multiple choice task where participants were instructed to mark the definition of the word, phrase or idiom they thought were correct. There were 40 questions in total in the test and four alternatives to each question. The items selected for the test were retrieved from all four episodes, being relevant to the plot or jokes presented in the clips. The idea behind conducting this test was to measure whether word definition aptitude had any significant relationship with grammar, vocabulary as well as lexical recall ability. We wanted to view the interconnection of several language components to better understand the data produced in each respective test. This is after all a quantitative study, and with that, the more data we have at our disposal, the more we can come to know.

Subsequently to the word definition task, I conducted an electronic, E-prime built lexical recall task (the list of words used is shown as a list in appendix D), in which I measured whether the students would remember episode-specific lexical items. The lexical recall task was built with the help of the Department of Modern Foreign Languages in E-prime. In this way, we were able to measure both target accuracy and response time. The test itself consisted of 60 words, varying in frequency and word type (noun, verb and adjective). 40 of these words occurred in the viewed episodes, while the remaining 20 did not. The largest challenge in choosing what lexical items to use in the digital testing was finding words of medium to high frequency that passed certain criteria that we decided upon. The first criterion was that the words chosen needed to be relevant to the plot or gags of the four episodes. In that way, the participants should be able to pick up on these lexical items more easily. The second criterion was the most difficult one: the occurring words should not be words that the students would frequently come across outside the episodes in the classroom, or which they might already be too familiar with. For example, the word “curriculum” is according to the corpus of contemporary American English a high frequent word, but would in this event not be included because the students would have come across it in English class when learning about the teaching plan. However, we suspected the students would be more prone to remembering low frequency words as opposed to medium or high frequent words and connect them with the episodes, because of the fact that they would less likely have come across the word other places. The occurring low frequency noun “paraplegic” is relevant to the plot of episode 14 in the fifth season of Family Guy, both because the character Joe is disabled and because he is denied access to a restaurant as a consequence of his disability. It is therefore more feasible to believe the participants would recognize the word in relation to the episode. This is one of many examples of words chosen for the lexical recall task.

The test itself was in many ways similar to the vocabulary test. One word would occur at a time on a computer screen, and the participants would enter a key for ‘yes’ or ‘no’ for whether they had come across that lexical item when watching the episodes. We had labeled the respective keyboard keys clearly in advance. We chose to use two laptops and testing two students at once, because we could then counter-balance the test by changing the formation of input keys on both laptops. One laptop used the ‘Z’ key for ‘yes’ and the ‘M’ key for ‘no’, and the opposite with the other laptop. Before starting the test, the subjects entered their information for participant identification in various boxes on the computer before a short, but informative text would explain the basics of the test. I explained the test thoroughly to the

class in plenum before actually conducting the test, but the information on the screen was added just in case. We placed the laptops used for the experiments towards the wall so that the participants would not be distracted by events occurring outside. The students were not disturbed during the testing to avoid influencing their performance, but we were still present in the room.

The results of the digital test were collected, merged and analyzed in R. The results chapter will report and further elaborate on the actual findings. When reading the raw data from the responses, I came across some significant outliers, namely response times that were too long. One instance of this was a word that had a response time of 40 000ms (40 seconds), which was discarded from the results. Most responses ranged from 500-800ms, with no responses that were suspiciously fast (less than 200ms).

3.5 Analysis

The data gathered from all respective tests were generated for the purpose of quantitative measurements, in other words a statistical analysis. All data relevant for this study was input in Microsoft Excel in a list format in order for it to be compatible with the program R. R was used in order to analyze, check for outliers, compare results of all three groups involved, output several different visual graphs. Furthermore, it was important to scrutinize effects of manipulated variables on dependent variables such as reaction time and accuracy, using analytical tests such as Analysis of variance (ANOVA), Shapiro-Wilk test for normality distribution and a linear mixed model. If there is a significant shift in lexical recall score within one of the three groups or even within the same individual subject, the background information questionnaire, grammar test or vocabulary test scores might be significantly correlated in some way. Nevertheless, the processed data will not elaborate much in itself as raw numbers and graphs, but will be comprehensible for a discourse in the discussion chapter. The context of the discussion will inquire and inspect the statistical data in order to find out whether the potential difference variable values between groups and within subjects proves or falsifies the hypothesis of this thesis. The results of all relevant statistical analyses will be shown as boxplots. This study tested a small sample of participants in each of the experimental groups. Therefore the results from the analyses of the data should be seen as exploratory and as revealing some tendencies.

4.0 Results

The data from all respective tests were structured in Excel and subsequently input in R for analysis. Some data had to be excluded for various reasons. This includes extreme values in the digital testing or funny or joke responses (although these were rare), and even whole participants. In total, 3 participants were excluded from the project because of underexposure to the film material, and one of these subjects had reading/learning difficulties, which could compromise the integrity of the testing. Therefore, results from 21 participants were processed in R. One should keep in mind that a test of this latitude would ideally inquire a broader amount of participants, meaning that readers will have to acknowledge the representativeness of the populace used in this experiment.

4.1 vocabulary and grammar tests

Table 1 | Scores from background vocabulary and grammar tests. Average scores, standard deviations, and normality distributions from the Shapiro-Wilk test

	EG	NG	CG
Vocabulary Avg.	37.28571	48.42857	40.14286
Grammar Avg.	30.57143	37.71429	34.28571
Stdev Voc	11.996031	14.128661	11.781745
Stdev Gram.	4.7908643	5.3763149	4.9665548
P-value Voc.	0.5006	0.3038	0.02308
P value Gram.	0.7174	0.9079	0.7715
W-value Voc	0.92393,	0.89536	0.75656
W-value Gram	0.94865	0.97134	0.95388

EG = English subtitles group | NG = Norwegian subtitles group | CG = Control group
Avg = average score | Voc = Vocabulary test | Gram = Grammar test | Stdev = standard deviation

The scores of each individual group was averaged in excel, but analyzed with R to view normality distribution using a Shapiro-Wilk test. The discussion chapter will elaborate further on these numbers in relation to the scores of the digital test that was built in E-

prime, but one should take a note of the ‘Norwegian subtitles’ group’s score. We can see that participants in this group performed generally better than its counterparts in both tests, but most notably with the vocabulary test.

The maximum score of the vocabulary test was 100, as it was measured in percentage, while the grammar test had a maximum score of 50. The vocabulary test scores as a whole output a p-value of 0.1473 when applied with a Shapiro-Wilk test while the grammar test gave a p-value of 0.5326. An interesting observation is that the control group performed better than the English group at the grammar test, but was out-performed by the English group in the vocabulary test. A linear mixed model was conducted in order to see whether grammar and vocabulary had an effect over each other, as is seen in table 2. Furthermore, the independent variable Group (each of the three groups in the experiment) was added into the analysis to see whether group type had an effect over grammar and vocabulary scores.

Table 2 | linear mixed mode of vocabulary and grammar test scores, and all three groups

	Estimate Std	T-value	Pr(> t)
Voc~Gram	0.7004	5.334	2.10e-07 ***
Gram~Voc	0.14181	5.334	2.1e-07 ***
Voc~Group	2.8378	2.999	0.00297 **
Gram~Group	1.0502	2.453	0.0148 *

Voc = Vocabulary | Gram = Grammar | Group = All individual groups | estimate std = estimate standard deviation

Table 2 indicates that the vocabulary and grammar skills of the participants were strongly correlated, as measured on the two on-line tests. Furthermore, we see that the independent variable Group had an effect on both Vocabulary and Grammar performance ($p < 0.05$), indicating that the three groups differed significantly on the L2 measures on the outset.

4.2. Word definition task

Figure 1 | Boxplot of word definition task scores (Y-axis) of all three groups (X-axis)

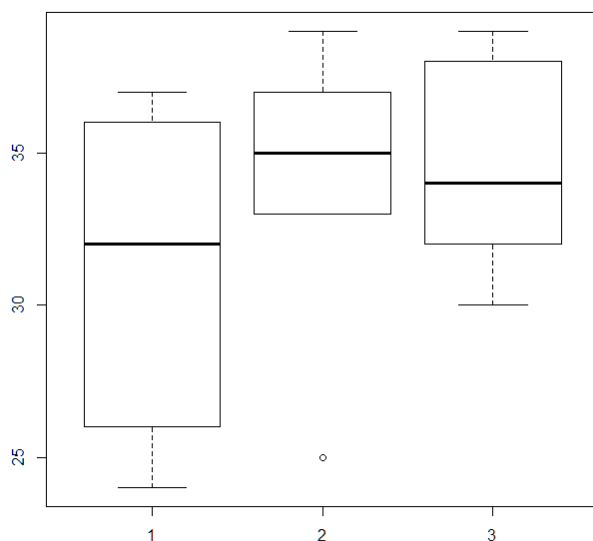


Figure 1 shows the average scores of the three individual groups. All three group scores were high, as was expected due to the task being a multiple choice task with one correct answer, two funny/odd, but wrong answers, and one completely false. The maximum score was set at 40, as each question rewarded 1 point.

An immediate observation of figure 1 illustrates a rather large discrepancy between the results of the English subtitles group and the two other groups. Considering that the English subtitles group performed considerably lower on both grammar, vocabulary and subsequently the word definition task, it becomes all the more important to view this fact when discussing the results of the lexical recall task, since there is a clear difference in English proficiency between the three groups. The boxplot suggests that the English subtitles group is far less proficient with definitions of English words, phrases or idioms, something that puts their overall English proficiency into question. The control group, having viewed the material without subtitles, performed relatively evenly with the Norwegian subtitles group. In fact, the scores of the word definition task seem to mirror the scores of the vocabulary and grammar tests, which will therefore be looked into further.

Table 3 | Analysis of the scores on word definition task | average score, standard deviation and Shapiro-Wilk test

	Avg.	Stdev	W-value	P-value
EG	31.57143	4.894117	0.90633	0.3711
NG	34.14286	4.598136	0.8691	0.1823
CG	34.5	3.674235	0.90689	0.4163

EG = English subtitles group | NG = Norwegian subtitles group | CG = Control group

The scores were revealed to be rather uniform, which immediately shows no large margin of difference between groups. However, since the three groups performed largely different on the vocabulary and grammar tests, it became necessary to construct a linear mixed model to measure whether the scores of those two tests could predict the results of the word definition task, shown in table 4:

Table 4 | results of the linear mixed model of the word definition task and vocabulary, grammar and group

	Estimated Stdev	T-value	Pr(> t)
Wdtask~Gram	0.17219	10.183	< 2e-16 ***
Wdtask~Voc	0.24762	6.589	2.49e-10 ***
Wdtask~Group	1.4942	4.659	5.09e-06 ***

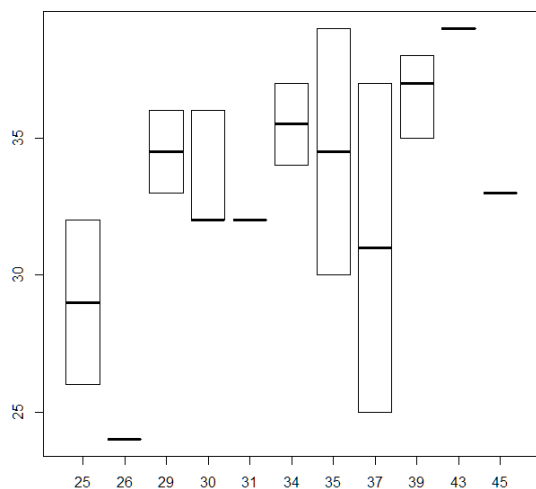
Wdtask = Word definition task | >0 = *** | >0.001 = ** | >0.01 = * |

The model reveals that both grammar and vocabulary scores predict the results of the word definition task, shown by the small P-value and large T-value. While the scores of the word definition task were rather similar between groups, it became crucial to measure the respective groups in relation to their vocabulary and grammar test performance. The fact that the scores of the word definition task seemed to mirror the results of the grammar and vocabulary tests made it a priority to look into what looked too suspicious to be a coincidence. The variable Group was also added in order to measure whether difference in group could

predict the scores of the word definition task, something the low P-value ($p < 0.05$) shows, indicating that variance of group does indeed predict the word definition task results.

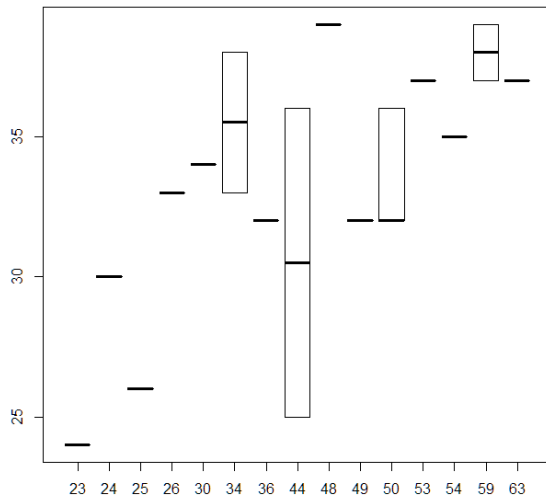
Whether the factor of subtitling is the leading variable predicting these results will be reflected upon in the discussion chapter.

Figure 2 | Boxplot of word definition task (Y-axis) ~ Grammar test scores (X-axis)



An observation of the output in figure 2 tells us that higher scores on the grammar test mean higher performance on the word definition task. Note that some participants reached the same score on an individual task, and are thus counted as the same degree in the boxplot. This goes for both figures 2 and 3.

Figure 3 | Boxplot of word definition task scores (Y-axis) ~ Vocabulary test scores (X-axis)



In concord with figure 2, figure 3 reveals that higher performance on the vocabulary tests implies higher performance on the word definition task. Vocabulary seems to have the largest impact of the two, which is natural due to word definition aptitude being closer to vocabulary than grammar.

4.3 Lexical recall test

Table 5 | Results of Accuracy (Acc) and reaction time (RT) on the lexical recall task | average scores and results of the Shapiro Wilk test

	Acc avg.	RT avg.	W-value Acc	W-value RT	P-value Acc	P-value RT
EG	0.557346	1652.315	0.8953	0.87006	2.25e-06	2.057e-07
NG	0.547927	1642.488	0.91636	0.74157	2.143e-05	2.338e-11
CG	0.539239	1502.312	0.92538	0.88111	1.207e-05	2.711e-06

EG = English subtitles group | NG = Norwegian subtitles group | CG = Control group

This was the final test conducted in the series, and the most essential one at that. Accuracy was measured with 1 point for correct and 0 for incorrect answer, thus the scores were estimated between 1 and 0. Reaction time was measured in milliseconds (ms) where extreme values, though few, were excluded. We immediately see that the English subtitles group

performed better on Accuracy and had almost the same average of reaction time as the Norwegian subtitle group, whereas the control group responded quicker, yet scored lower than its counterparts on accuracy. Note that the P-values for all groups on both Acc and RT are considerably low.

Figure 5 | Boxplot of lexical recall task scores (Y-axis) of all three groups (X-axis)

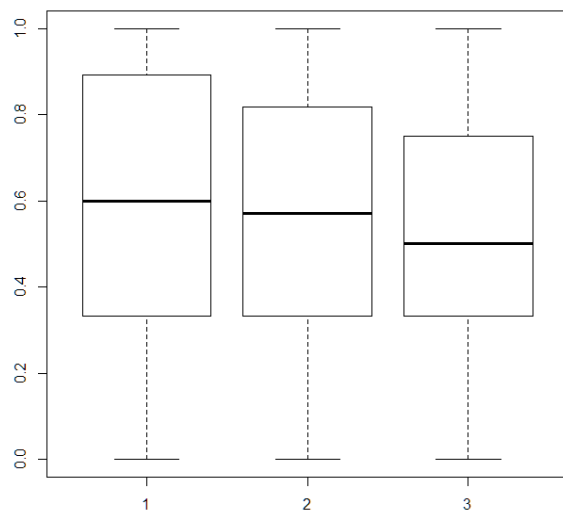


Figure 5 shows that the English subtitles group had a higher mean score compared to the two other groups. The lower quartile for all groups was almost identical, but the higher quartile was different between all groups, in that the English group had overall more correct answers than the Norwegian group, and lastly the control group at the bottom. The fact that the English subtitles group outperformed the two other groups in lexical recall, when they in return outperformed the English group in vocabulary, grammar and word definition, creates a strong argument for intralingual subtitling being a beneficial agent in their performance. However, this is just an immediate observation, requiring us to analyze the data output to fully understand the difference in performance between groups.

Table 6 | Analysis of variance results of the lexical recall task

	F-value	Pr(>F)
Accuracy~Freq	F(2,40)= 7.756	0.00142 **
Accuracy~Wclass	F(2,40)= 0.158	0.854
Accuracy~Occur	F(1,20)= 0.103	0.752
RT~Freq	F(2,40)= 1.404	0.257
RT~Wclass	F(2,40)= 5.852	0.0059 **
RT~Occur	F(1,20)= 0.811	0.378
Accuracy~Group	F(2,270)= 1.132	0.324
RT~Group	F(2,270)= 1.132	0.324

Freq = Frequency | Wclass = Word class | Occur = Occurance | instances with the variable ‘Group’ is an between subjects model – the rest is a within subjects model

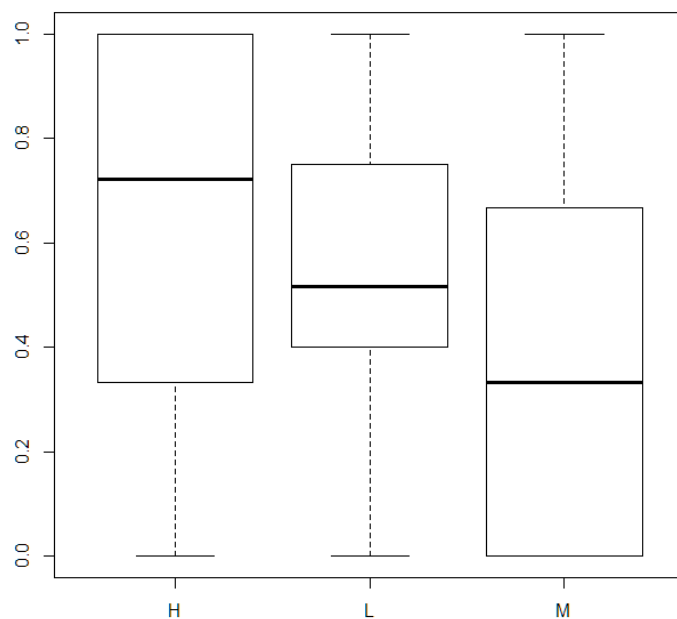
>0 = *** | >0.001 = ** | >0.01 = * |

The first variable before the tilde is the dependent variable; the other is and independent variable.

An anova (analysis of variance) test was necessary here in order to find out whether an independent variable had an effect on a dependent variable. The independent variable Group contained the three individual groups used for testing, and was treated in R as such when scrutinizing for evidence whether there was an effect on either Accuracy or RT. The anova test reported that there was no significant effect of Group on either Accuracy or RT. Freq appeared to have an effect on accuracy, namely in that participants scored higher on high frequent words as opposed to low and medium. This effect has been drawn as a boxplot in figure 6. Wclass had a significant effect on RT, which is a common instance in psycholinguistics – subjects tend to respond quicker on content words, as is shown in figure 3.

A high F value in both instances indicates an effect, while a low P value indicates that the null hypothesis can be discarded in this analysis, showing clear evidence that there is an effect.

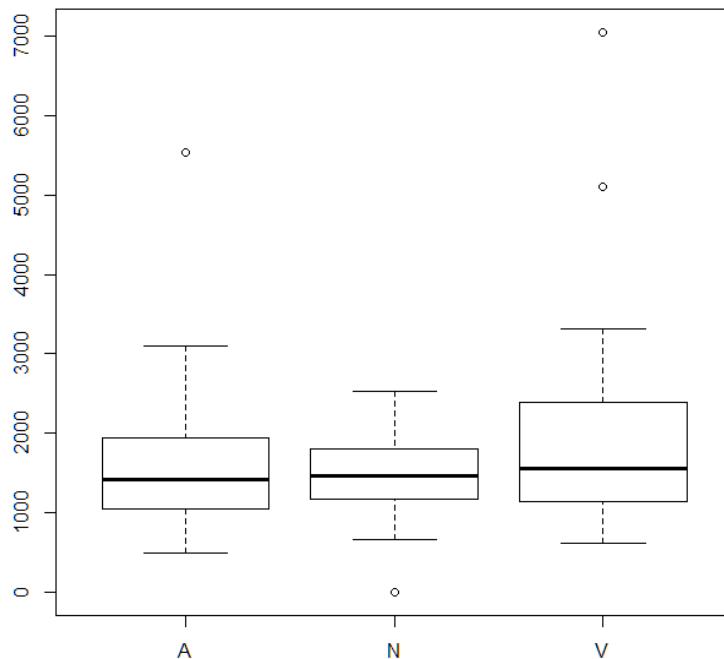
Figure 6 | boxplot of accuracy scores ~ word frequency, both from the lexical recall task



Y-axis = Binary score of 0-1 (participants received either 1 for correct input and 0 for wrong)
X-axis = Frequency: High, Low and Medium

Figure 6 shows that overall Frequency has an effect over Accuracy, as we can see that the participants scored higher on high frequency lexical items, but relatively lower on medium frequency as opposed to low frequency words. However, as distinguished earlier, medium and low frequent words will be treated the same, since they were relatively close to one another in terms of score in the corpus of contemporary American English (COCA). A score of 1-4999 was regarded as low, 5000-9999 medium and all above 10 000 as high frequency.

Figure 7 | Boxplot of reaction time ~ word class



X-axis = Word class (Adjective, Noun and Verb)

Y-axis = Reaction time (reported in milliseconds)

Figure 7 shows that participants responded faster when encountering adjectives and nouns as opposed to verbs, which is a common instance in psycholinguistics where subjects process nouns relatively faster than other word classes. Content words are easier to recognize and faster to process for the brain. This finding is nothing new, but shows that the data adheres to conventional psycholinguistic theories, meaning they signify their validity for that reason. For the sake of proper data analysis, extreme values were excluded, being reaction times surpassing 10000ms. Participants were expected to be able to respond well within 2000ms, but would frequently wait a few seconds, which one can assume was to actively recall whether they had encountered the particular word or not. Therefore, a threshold of 10000ms was established, as outlying numbers appear suspicious (i.e. participants becoming distracted from the test for various reasons).

Table 7 | Linear mixed model of accuracy scores on lexical recall task ~ vocabulary test scores + grammar test scores (Accuracy~Voc+Gram)

	Estimate	Stdev	T-value	Pr(> t)
Voc	0.003441		-1.951	0.0521 .
Gram	0.003482		0.889	0.3751

>0 = *** | >0.001 = ** | >0.01 = * | Voc = vocabulary test scores | Gram = grammar test scores

Table 8 | Linear mixed model of reaction time scores on the lexical recall task ~ vocabulary test scores + grammar test scores (RT ~ Voc + Gram)

	Estimate	Stdev	T-value	Pr(> t)
Voc	-4.244		-1.071	0.28497
Gram	24.482		2.781	0.00582 **

>0 = *** | >0.001 = ** | >0.01 = * | Voc = vocabulary test scores | Gram = grammar test scores

Vocabulary shows a borderline effect on Accuracy in table 3 (Accuracy ~ Voc + Gram), revealed by the P-value 0.0521, which is very close to the standard 0.5 threshold. This requires attention, as it does not mean we can automatically reject the null hypothesis, but since it is so close to the threshold, it requires special attention. For the time being, we only have evidence that suggest that there might be an effect, but it is not solid enough evidence to discard the null hypothesis. However, the scores from the grammar test are revealed in this model to predict RT (reaction time), which is shown in figure 4.

5. Discussion

5.1 Vocabulary and grammar

Testing for vocabulary and grammar was conducted in order to view the participants' English proficiency based on these two language components. The importance of these measurements became most relevant when the results revealed a large discrepancy of scores between groups, something that was paid special attention to. Both grammar and vocabulary are important factors in a language, thus influencing performance on other language tasks. The English subtitles group scored on average 37.28 %, while the Norwegian subtitles group reached 48.42 % and lastly the control group with 40.14 %. The contrast is clearest between the Norwegian and English subtitles group with a difference as large as of 11,14 %. On the grammar test, the English subtitles group got on average 30.57 (out of the total of 50), the Norwegian group 37.71 and the control group 34.28. The Norwegian subtitles group outperformed both other groups on both tests, creating an expectation that they would perform relatively higher on word definition and lexical recall due to their higher proficiency in English. The English subtitles group were expected to perform lower on subsequent tests due to their assumed lower proficiency with the English language based on results from both tests.

5.2 Word definition task

The results from the word definition task show at a first glance that subtitles did not have any perceivable effect on the scores produced by each individual group. All groups performed relatively uniform, with the Norwegian subtitles group performing best of all three. However, one must recall the scores of the grammar and vocabulary test, where the Norwegian subtitles group outperformed both the English subtitle group and the control group by a large margin, making them expected to perform better at a word definition task. This task was constructed in order to measure whether subtitles had significant effects on definitions of lexical items and idioms that were observable in the viewed episodes. A linear mixed model in R showed that both grammar and vocabulary test scores predicted the results of the word definition task, meaning that their skills in grammar and vocabulary predicates their word definition competence. The lexical items, including singular words, idioms and other phrases in the word definition task were taken directly from all four

episodes, which all participants had viewed. However, the fact that they had encountered these items does not itself predicate their knowledge of definitions. As Nagy and Scott puts it, “[...] knowing a word cannot be identified with knowing a definition. (Nagy and Scott, 2009, p. 273). With this in mind, one cannot assume that the participants would naturally be able to answer correctly on the items’ definitions just because they have been exposed to them. Nevertheless, many of these items could relatively easily be defined by the subjects by understanding the context of the terms, words and phrases. The question is whether they remember these instances from the episodes and additionally know the definitions of them. For example, the word “ordeal” was uttered by a character from the series in the sentence “Boy, that was an ordeal!” which follows immediately after the character Peter has to sit through a very long and tedious television show’s opening song. From the context, one can easily assume that the definition “A test of patience” is the most correct answer in contrast to “A test of physical strength”, “A drug test” and “A brand of chewing tobacco” if the participant remembers the scene in combination with this word. More of examples can be found in appendix C.

The fact that there is evidence to suggest that grammar and vocabulary proficiency predicates word definition task performance, does not by itself suggest that subtitles helped them score higher. The results show that the control group performed relatively better than the two other groups, by a small margin. Viewed in relation to scores on grammar and vocabulary, we see a different outcome. Kvitnes (2013) reported that her Norwegian subtitles group performed better than the two other groups, and similar to the results in my experiment, they scored higher on grammar and vocabulary than the contrasting groups. She argued that Norwegian subtitles could have helped their performance, something not replicated in this study. Here, the output seems to suggest that the control group, without any subtitles, performed better than the Norwegian subtitles group which interestingly enough performed much better on both grammar and vocabulary. There are of course many distinguishable factors that could affect the dataset we have generated, such as participant attention, something Kvitnes also projects as a possible weakness of the design (ibid). The students could also be familiar with the definitions beforehand, unrelated to the episodes, and could have taken lucky guesses. In any case, the word definition task cannot by itself prove whether subtitles are beneficial for this particular task, purely because other factors can override the projections. The questions were also relatively easy, as it was designed for their skill level (VG3) to be able to handle well if they had paid attention to the respective

episodes. Nevertheless, the interconnection of vocabulary and definition is precisely what Zimmerman (1997) argues when discussing the historical aspects of language learning (p. 6 and 7). In hindsight, it would have been rather interesting to see whether a potential second control group, which would not have been exposed to any of the episodes at all, would be able to score as high as the groups that did watch the episodes. That way, one could measure whether audiovisual second language stimuli would predict word definition aptitude.

Since there is strong evidence in the data output that high vocabulary and grammar test performance indicate high word definition task performance, it means that we have to look further into this correlation. Another linear mixed model which measured whether the difference in group and their respective scores on the vocabulary and grammar tests had an effect on their performance on the word definition task was constructed. From this, we see that there is indeed strong evidence for significance. A boxplot (chapter 4.2. figure 1) showing the three groups' performance on the word definition task reveals that the English subtitles group had a data cluster considerably lower than the other two groups, although having a mean value much closer to its counterparts. Accepting that the English subtitles group performed lower than the two other groups on the vocabulary, grammar and as well as the word definition task, we must acknowledge their low performance as an impact on their English proficiency, thus deeming their overall English considerably lower than the two other groups. This becomes rather important when looking at the results of the lexical recall task, which measures primarily accuracy on word recall ability.

5.3 Lexical recall task

As table 5 in chapter 4.3 shows, the all groups performed relatively evenly, with the English subtitles group coming out on the top. With the argument made in chapter 5.1 that the English subtitles group underperformed in terms of grammar, vocabulary and the word definition task, it becomes considerably interesting that the group outperformed the two other in the lexical recall task. This task measured both Accuracy and reaction time, with a main focus on the former. The idea of the test was to see whether participants could distinguish occurring from non-occurring lexical items in the form of singular verbs, nouns and adjectives, from the four episodes of Family Guy. Note that in the discussion of

frequency, low and medium frequent words will be regarded as the same, since there is little to distinguish one from another apart from frequency score from the Corpus of Contemporary American (COCA). Thus, I will only refer to two frequencies: high and low.

An analysis of variance (anova), shows that word frequency had an effect on the dependent variable Accuracy. All subjects of all three groups scored higher when encountering high frequent words, occurring or non-occurring, as shown in figure 6. One must recall that the participants were tested well over a month after finishing viewing the episodes, meaning that in a general sense, higher frequent words were more easily recollected than lower frequent words over a long period of time. The results therefore signify that word frequency is one of the long-term effects on learners' language proficiency through watching audiovisual material. However, the main focus of this study is to measure the impact of subtitles, meaning that the difference of accuracy between all three groups is the key factor to responding to the research question at hand. At a first glance, the anova results in table 6 show that there is no significant effect of Group on Accuracy, meaning that each group could not predict the scores on accuracy. However, the argument made in chapter 5.1 about the discrepancy of grammar, vocabulary and word definition task scores between groups made it a priority to measure whether vocabulary and grammar could predict accuracy on the lexical recall task. A linear mixed model revealed a borderline effect (a P-value of 0.0523) of vocabulary on accuracy, which by convention means that we need to pay special attention to it. A borderline effect suggests evidence against the null hypothesis, but since it is not beneath the threshold; one cannot completely discard it. What it does, though, is signaling that vocabulary might be a predictor for lexical recall aptitude. If we revisit the fact that the English subtitles group performed considerably lower on vocabulary than its two counterparts, and at the same time consider that vocabulary test scores have a borderline effect on word accuracy, it becomes all the more possible that the addition of intralingual subtitles for the English subtitles group benefitted their performance on the lexical recall task. This therefore becomes a strong argument for there being a beneficial long-term effect of using subtitles, particularly intralingual, when viewing audiovisual material. Martine Danan (2004) points to various studies that implicate positive vocabulary effects of using intralingual subtitles, writing that subtitles have:

“[...]demonstrated the positive effects of captioning on productive skills such as verbatim recall and retention, reuse of vocabulary in the proper context, as well as communicative

performance in specific oral and written communication tasks” (Danan, 2004, p. 69, reference to Vanderplank, 1988)

With the results from the lexical recall task in mind, the argument made by Danan fits appropriately with the marginal effect between vocabulary and lexical recall accuracy. Both the Norwegian subtitles group and the control group performed better on the vocabulary, grammar and word definition tasks, but were outperformed by the English subtitles group on accuracy on the lexical recall task. An expectation therefore arises, that if all groups performed precisely uniform on vocabulary, grammar and word definition, the English subtitles group would have performed considerably higher on lexical recall accuracy because of the interconnection of vocabulary with lexical recall, as Zimmerman sustains (1997). This naturally requires us to call for more research on the same topic with a greater sample of the population for a more ideal measurement. The borderline effect presented in the linear mixed model signals a potential effect, meaning that the special attention required necessitates further scrutiny of the effect. As of now, the data implies that viewing L2 subtitles with L2 video positively affects learners’ vocabulary, in line with Koolstra’s (1999) panel study on second language subtitles. Grammar remains more or less unaltered.

A linear mixed model showed that the scores of the grammar test revealed to have an effect on reaction time (RT) in the lexical recall task. In other words, higher performance on grammar indicates slower time between seeing the word appear on the screen and pressing either yes or no. Reaction time in within-subjects (within all groups) contribute little to an idea that grammar differences predict reaction time apart from signaling that participants with lower grammar scores spent less time on each word. A reaction time experiment on various age groups by Hulstsch (2001) showed that Individual differences in RT inconsistency correlated negatively with level of performance on measures of perceptual speed, working memory, episodic memory, and crystallized abilities. A slow reaction time does not necessarily indicate a negative effect, considering participants could have spent longer time trying to actively recall whether they had encountered the particular word. The impact that higher grammar could have on reaction time, is that students with higher proficiency with grammar could be more concerned with the quality of their performance rather than affectivity. The students were not told that they would be measured on reaction time, meaning that they would not know this would be a factor in the dataset. A consequence of that is that the participants would spend as much time as they personally

felt was needed on every response, generating more natural data than if they were aware of the concept of reaction time as a measurement.

5.5 Limitations of the study

An initial, perhaps rather aspirational goal of this study was to expose the participants to even more stimuli (a whole season of *Family Guy*). This had to be limited down to four episodes over four weeks due to the tight schedules of the students' study programs. Ideally, we wanted to show almost an entire season (over twenty episodes) of the television series over the course of four weeks, but the constriction of four episodes did however enable us gather material for the lexical recall task and the word definition task more efficiently. The number of lexical items used for either test would effectively be the same, but with more episodes in the scope of exposure, we would be able to cherry-pick items from a wider range of episodes, thus providing both a more ideal counter-balance and a more diverse set of words, phrases and idioms for the respective tests.

Another limitation of this study is the sample population for the experiment, being that only 21 participants were able to participate. Initially, we wanted to replicate and even expand the amount of participants used in the study of Vulchanova et. al, to generate a diverse amount of data for analysis, but only one class was able to participate. As a consequence, the data output had to be scrutinized particularly cautiously, and any test had to be counter-balanced and fitted appropriately wherever possible in order to generate the most naturalistic data conceivable.

6. Conclusion

This study was an attempt to scrutinize the potential long-term impacts subtitled audiovisual material has on second language acquisition. We investigated this through a quantitative and experimental research approach, where twenty-one 18-year-olds partook in various language proficiency assessments, and were then divided into experimental groups and exposed to stimuli in form of four episodes of *Family Guy* over the course of four weeks. The first group watched the episodes with intralingual (English) subtitles, while the second one watched with interlingual (Norwegian) subtitles. The third group acted as a control group, and watched episodes without any subtitles. To measure the potential long-term effects, we waited approximately one month after the stimuli to test the participants on word definition and lexical recall. Our data showed that both vocabulary and grammar predicted word definition task scores on each group, indicating that higher proficiency in grammar and vocabulary lead to better scores on the word definition task. The English subtitles group performed considerably lower than the other two groups in terms of grammar, vocabulary and word definition, deeming them to be less proficient in English than its counterparts. While the English subtitles group performed considerably lower than the other groups on grammar, vocabulary and word definitions, they outperformed the other two groups on accuracy on the lexical recall task. This result suggests that this group had actually benefitted from exposure to the intralingual (English) subtitles.

The studies by Mitterer & McQueen (2009) as well as Vulchanova et al. (2013) illustrate that subtitled audiovisual material may facilitate second language acquisition, something which Baltova (1999) affirms. The results of this study are consistent with previous research, documenting the importance of various aspects of language such as frequency in lexical recall and retrieval. We also showed that word class plays a role in lexical recall. The most significant attainment has been showing that intralingual (L2 to L2) subtitling had the most significant impact on lexical recall aptitude, supporting that it was beneficial for their second language acquisition. Long-term effects as a consequence of audiovisual subtitled material has had considerably little study in the larger scope of psycholinguistics, and this project has barely scraped the surface of what can be learned about the topic. Aurstad and Kvitnes (Vulchanova et al., 2013), called for more research and study on how

subtitles implicate effects on second language acquisition. Having done just that, I would be more than obliged to encourage further study on the same topic.

Works cited

Borrás, I., Lafayette, R. C. (1994). Effects of Multimedia Courseware Subtitling on the Speaking Performance of College Students of French, *The Modern Language Journal*, 78(1), (p.61). Doi:10.2307/329253

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

Swaffar, J., (1999). Review, *The Modern Language Journal*, 83(3), (pp. 440-441)

Plass, J. L., Jones, L. C. (2005). Multimedia Learning in Second Language Acquisition. In Mayer, R. E. (ed.), *The Cambridge Handbook of Multimedia Learning*, (p. 467). UK: Cambridge University Press

Rammeverk for grunnleggende ferdigheter. (n.d.). Retrieved on 29. April 2016 from Udir: http://www.udir.no/Upload/larerplaner/lareplangrupper/RAMMEVERK_grf_2012.pdf?epslanguage=no

Saville-Troike, M. (2006). *Introducing Second Language Acquisition*,(p.2) UK: Cambridge University Press,

Cook, V. (2013). *Second Language Learning and Language Teaching Fourth Edition*. (pp. 6-7), London: Routledge

Johnson, M. (2004). *A Philosophy of Second Language Acquisition*)

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, (p.1-10). doi: 10.3389/fpsyg.2014.01510

Johnson, M., (2004). *A Philosophy of Second Language Acquisition*, (pp. 18-29), Yale: Yale University Press

Nornes, A. M. (1999). For an Abusive Subtitling, *Film Quarterly*, 52(3), 17-34. DOI: 10.2307/1213822

Turek, S. (2010), Foreigners in the Margins: English Subtitles in “Inch’Allah Dimanche”, *The French Review*, 83(3), 560-572

Koolstra, C. M., Beentjes, J. W. J., (1999). Children Vocabulary Acquisition in a Foreign language Through Watching Subtitled Television Programs at Home, *Educational Technology Research and Development*, 47(1) (pp. 51-60). Springer.

Gass, S. M., MacKey, A., Pica, T., (1998). The Role of Input and Interaction in Second Language Acquisition: Introduction to the Special Issue, *The Modern Language Journal*, 82(1). (pp. 299-307), National Federation of Modern Language Teachers Association

Lee, J. F., Benati, A. G., (2009). *Research and Perspectives on Processing Instruction*, (p. 4), Berlin: Mouton de Greyter,

Verspoor, M., Lowie, W., de Bot, K. (2007), *Input and Second Language Development from a Dynamic Perspective*, (p. 1.) Department of Applied Linguistics, University of Groningen

Baltova. I, (1999) The Effect of Subtitled and Staged Video Input on the Retention of Content and Vocabulary in a Second Language (p.1.), University of Toronto: Toronto

Blystad, E. S., Maasoe, A., (2004) *The Invisible Text*, (p. 6), Oslo; University of Oslo

Munday, J., (2012), *Introducing Translation Studies, 3d edition*, (pp. 271, 278-279) London: Routledge

Zimmerman, C. B., (1997), Historical trends in second language vocabulary instruction, in Coady, J. Huckin, T (eds.), *Second Language Vocabulary Acquisition A Rationale for Pedagogy*, (pp. 5-7), London: Cambridge University Press

Nagy, W. E., Scott, J. A., (2009) Vocabulary Processes, in Kamil, M. L., Mosenthal, P. B., Pearson, P. D., Barr, R., (eds.), *Handbook of Reading Research Volume III*, (pp. 269-273). London: Routledge

Hultsch, D. F., MacDonald. S. W. S., Dixon, R. A., (2001), *Variability in Reaction Time Performances of Younger and Older Adults*, doi: 10.1093/geronb/57.2.P101 British Columbia: University of Victoria

Zhang, D., (2012), Vocabulary and Grammar Knowledge in Second Language Reading Comprehension: A Structural Equation Modeling Study, *The Modern Language Journal*, 96(4), 558-575, National Federation of Modern Language Teachers Association

López-Burton, N., (2014), *On Being a Language Teacher: A Personal and Practical Guide to Success*, (p. 24.), Yale: Yale University Press

Foster-Cohen, S. H., (1999), *SLA and First Language Acquisition*, (p.4), USA: Cambridge University Press. 10.1017.S0267190599190019

Aurstad, L. M. G., (2013), *The Role of Subtitles in Second Language Acquisition*, NTNU, retrieved on 29. April, 2016,
https://www.ntnu.edu/documents/38274309/40279009/LisaMGAurstad_Master_2013_withtitlepage.pdf/190bc71d-3e85-4519-a78b-5064597e5298

Kvitnes, I. I. N., (2013), *Subtitles in the Second Language Classroom*, NTNU, retrieved on 29. April, 2016,
<https://www.ntnu.edu/documents/38274309/40279009/Ingrid+Elisabeth+Nufsfjord+Kvitnes+Master.pdf/5d0bb5fe-b010-479e-90e2-5939766aa2ac>

Utdanningsdirektoratet, (2006), *English – Programme Subject in Programmes for Specialization in General Studies*, retrieved on 11. May, 2016, Udir,
[http://www.udir.no/Stottemeny/English/Curriculum-in-English/Upper-secondary-education-/](http://www.udir.no/Stottemeny/English/Curriculum-in-English/Upper-secondary-education/)

Froelich, J., (1988), *Die Unterrichtspraxis / Teaching German, German Videos with German Subtitles: A New Approach to Listening Comprehension Development*, 21(2), (p. 199). DOI: 10.2307/3530283

Appendices

Appendix A – Consent form

Samtykke til deltakelse i undersøkelse om andrespråkforståelse

Ansvarlig institusjon: NTNU.

Student: Henrik Eye (henrikey@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 november/desember, 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 Henrik ved eventuelle spørsmål.

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

Trondheim, Dato: _____

Underskrift: _____.

Appendix B – Background information sheet

Bakgrunnsinformasjon for forskningsprosjekt om andrespråksforståelse

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

Informasjonen som du oppgir vil 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 informasjonen. 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.

Vi ber deg legge merke til at skjemaet har totalt 6 sider.

Henrik Eye / Erlend York

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 i uka Et par ganger i uka Av og til Aldri

Hvor ofte skriver du tekst på norsk?

- Hver dag Flere ganger i uka Et par ganger i uka Av og til Aldri

Engelsk og andre fremmedspråk

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

	Grunnleggende	Middels	Avansert	Flytende
Lesing				
Skriving				
Snakke				
Lytte				

Totalt				
--------	--	--	--	--

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? _____

Hvilke språk kan du utover morsmålet ditt og engelsk?

Språk	Nivå			
	Grunnleggende	Middels	Avansert	Flytende
Tysk				
Fransk				
Spansk				
-Angi språk				
-Angi språk				
-Angi språk				

Hvor ofte leser du tekster på engelsk?

- Hver dag Flere ganger i uka Et par ganger i uka Av og til Aldri

Hvor ofte skriver du tekster på engelsk?

- Hver dag Flere ganger i uka Et par ganger i uka Av og til Aldri

Hvor ofte lytter du til/hører du engelsk?

- Hver dag Flere ganger i uka Et par ganger i uka Av og til Aldri

Hvor ofte ser du på engelskspråklige serier/filmer?

- Hver dag Flere ganger i uka Et par ganger i uka Av og til Aldri

Når du ser på engelskspråklige filmer, hvilken av disse alternativene bruker du mest?

- Undertekst på norsk Undertekst på engelsk Ingen undertekst

Hvor ofte ser du på engelskspråklige tegneseriefilmer/serier?

- Hver dag Flere ganger i uka Et par ganger i uka Av og til Aldri

Har du sett tegneserien “Family Guy”?

- Ja Nei

Hvis Ja, i hvor stor grad? _____

Hvor ofte spiller du engelskspråklige data/TV-spill?

- Hver dag Flere ganger i uka Et par ganger i uka Av og til Aldri

Hvilke type spill spiller du? _____

Hvor mange timer cirka per dag? _____

Hvor mye TV ser du på hver dag?

- 7 timer eller mer 5-6 timer 3-4 timer 1-2 timer aldri eller nesten aldri

Del C: Andre faktorer i språklæring

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

Ja Nei

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

Ja Nei

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

Ja Nei

Har du, eller har du 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 C – Word definition task

Note: correct answers are marked with “X”

Participant code:

Word definition task

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

1. “Just browsing”

- a. Looking, but not buying. X
- b. Having an addictive relationship with shopping
- c. Searching for answers.
- d. Scratching your eyebrow.

2. “To fit in”

- a. To make sure there is enough room for oneself.
- b. To be excluded.
- c. To lose weight.
- d. To belong to a group, plan or situation. X

3. Refreshing

- a. An unpleasant feeling.
- b. A sensation of relief. X
- c. The haptic feedback of interacting with a mobile phone.
- d. Smelling awfully.

4. Fuzzy

- a. Hairy. x
- b. Itchy.
- c. Pointy.
- d. Annoying.

5. Vocal chords

- a. An organ that creates sounds and speech. X
- b. A musical instrument.
- c. A church hymn.
- d. An angry mob.

6. Playboy shoot.

- a. A recreational game where participants fire foam-based pellets from toy guns at each other.
- b. Volunteering at a food bank for starving children.
- c. Selling or renting weaponry to an underage citizen.
- d. Filming or taking pictures of women for an adult magazine X

7. Credit sequence

- a. A safe way to purchase goods and services.
- b. An illegal tax fraud.
- c. A medical service for people with early onset Alzheimer's.
- d. A part of a television show where the names of the creators are shown. X

8. Ordeal

- a. A test of patience. X
- b. A test of physical strength.
- c. A drug test.
- d. A brand of chewing tobacco.

9. Advertising

- a. Craving attention.

- b. Promoting a service or product. X
- c. Ignoring an individual.
- d. Pandering to a specific demography.

10. “Not going down easily”

- a. Charging the enemy head-on.
- b. Barricading oneself inside a safe area.
- c. Challenging someone to a duel.
- d. Putting up resistance, despite the risk of losing. X

11. Policy

- a. Having the attitude of a crooked cop.
- b. A set of rules and guidelines. X
- c. A slang term for a police officer.
- d. A list of things to purchase.

12. Ivory hunter

- a. An individual who kills elephants for their tusks. X
- b. A talent scout.
- c. An expert on grand pianos and other classical instruments.
- d. Someone who employ people into temporary positions.

13. Sea shanty

- a. A whale hunter from the Faroese Islands.
- b. An underwater creature, fabled to sink and destroy sea vessels that tread too near.
- c. A work song, sung by workers onboard ships. X
- d. Being completely silent at sea, with the exception of waves and wind.

14. A treat.

- a. A type of diplomatic agreement.
- b. Abusing someone physically.
- c. Something special and pleasant one gives to or receives from others. X
- d. Something that cannot be returned.

15. Possessed

- a. Having illegal substances in one's pocket.
- b. Having one's body used as a host by something otherworldly. X
- c. Containing traces of narcotics.
- d. Physical activity for the sake of bettering one's health.

16. Church service

- a. Attending a communal worship at a church. X
- b. Asking neighbors for donations for the church.
- c. Refurbishing a church as a part of voluntary work.
- d. Working for free.

17. Pseudoscience

- a. Attempting to conduct scientific work, without using true scientific methods. X
- b. A true form of conducting science that adheres to the norms of scientific practice.
- c. Peer-reviewed articles that have their claims validated by other scientists.
- d. Researching and experimenting on apes and primates.

18. Socially backward

- a. Not greeting someone with a formal handshake.
- b. Being inadequate at interacting with other people. X
- c. Standing in the wrong direction.
- d. Being absent-minded.

19. Stubborn

- a. Being rather persistent. x
- b. Being difficult to understand.
- c. Agreeing with someone.
- d. Having a rough day.

20. A red state

- a. A geographical area where the popular vote goes to the Republican Party. X
- b. An emotional state of constant despair.
- c. To blush out of embarrassment.
- d. A country with a history of communist ideals in politics.

21. Nutjob

- a. Slang for a sexual act between two people.
- b. Someone who hates squirrels and other rodents.
- c. A mentally deficient individual. X
- d. An occupation where the employee sells nuts from a stand.

22. A pursuit

- a. A type of clothing, usually rather expensive.
- b. A type of mint-flavored chocolate.
- c. A container for make-up accessories.
- d. A chase after someone or something. x

23. Salvation.

- a. To apply medication.
- b. Being saved or protected from harm, in a spiritual sense. x
- c. To read a verse from the Holy Bible.
- d. Being condemned to eternal despair.

24. Adultery

- a. Putting trust into someone else.
- b. Discontinuing being childish permanently.
- c. Being unfaithful to one's spouse. x
- d. Growing up faster than expected.

25. Bonding

- a. Being tied up with a rope.
- b. Imitating secret agents.
- c. To feel distant from another individual.
- d. To feel connected to another individual. X

26. Underfunded

- a. Having received sufficient money
- b. Having received less money than needed X
- c. Being sponsored by someone untrustworthy.
- d. Being selfish with one's money.

27. Raise hell

- a. To channel forth unholy creatures from the underworld.
- b. To be nice and courteous.
- c. To make a lot of trouble. X
- d. To say or do something one will later regret.

28. Founding fathers

- a. A detective bureau who specializes in finding lost children.
- b. A nickname for males in their 50s.
- c. A group of individuals who formed the declaration of independence. X
- d. A tribe of Native Americans who rebelled against the European settlers in the early 1600s.

29. Expelled

- a. Being an exemplary individual.
- b. Being reinstated at an institution.
- c. Being put in jail indefinitely.
- d. Being disallowed to attend something. X

30. Homecoming

- a. Returning from the war.
- b. Relaxing without anyone else present.
- c. A parade to celebrate diversity.
- d. An annual high school dance. X

31. Offender

- a. Someone who breaks the law.X
- b. Someone who insults people frequently.
- c. Someone who makes politically incorrect jokes.
- d. Someone who is a model citizen.

32. “Pull some strings”

- a. To use influence to get something done. X
- b. Using puppets to convey a message.
- c. Dragging someone along.
- d. Objecting an order from someone in authority.

33. “Rigging an election”

- a. Using illegal means to win. X
- b. Making sure you succeed.
- c. To validate and assess an election to make sure candidates operate lawfully.
- d. Setting up a booth for the individuals who count votes.

34. “Putting someone down.”

- a. To end someone’s life. X
- b. To place someone on the floor or ground.
- c. To disappoint someone.
- d. To put someone under a lot of stress.

35. Wig

- a. A native-American tent.
- b. Artificial hair that you wear on your head. X
- c. A thin stick.
- d. A rather long braid.

36. Pollution

- a. Contamination by waste or other harmful substances. X
- b. A natural process that plants and flowers undergo for procreation.
- c. A way of solving a problem.
- d. Cleaning the environment.

37. “That’s baloney!”

- a. “That’s an Italian sausage!”
- b. “That’s pretty clever!”
- c. “That’s not my problem!”
- d. “That’s complete rubbish!”. 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. “He doesn’t sit right with me”

- a. “I do not trust his intentions”. X
- b. “He is seated far away from me”.
- c. “He will be here shortly”.
- d. “I trust him very much”.

40. “I’m an absolute wreck”

- a. “I have broken down and cannot operate.”
- b. “I don’t feel very well.” X
- c. “I’m recovering from a traffic accident”
- d. “I feel very good about myself.”

Thanks for participating!

Appendix D – Lexical recall task list (used with the E-prime built testing)

Occurring = 1, non-occurring = 0

Word	Occurance	Frequency	Class
model	1	High	Noun
crucifix	1	Low	Noun
wheelchair	1	Medium	Noun
adopted	1	High	Verb
buckle	1	Low	Noun
marmalade	1	Low	Noun
puppy	1	High	Noun
sneakers	1	Low	Noun
queer	1	Low	Adjective
mule	1	Low	Noun
executed	1	Low	Verb
exorcism	1	Low	Noun
witness	1	High	Noun
disinfect	1	Low	Verb
miserable	1	Low	Adjective
bigfoot	1	Low	Noun
paraplegic	1	Low	Noun
youth	1	High	Noun
reverend	1	Medium	Noun
orphanage	1	Low	Noun
community	1	High	Noun
values	1	High	Noun
challenge	1	High	Noun
beverage	1	Low	Noun
persuasive	1	Low	Adjective
static	1	Medium	Adjective
ridiculous	1	Medium	Adjective
academy	1	High	Noun
powerful	1	High	Adjective
legend	1	Medium	Noun
campfire	1	Low	Noun
immoral	1	Low	Adjective
frosting	1	Low	Noun
decoy	1	Low	Noun
degree	1	High	Noun
munch	1	Low	Noun
retaliate	0	Low	Verb
tinker	0	Low	Verb
disagreed	0	High	Verb

clairvoyant	0	Low	Adjective
apprentice	0	Low	Noun
intermediate	0	Low	Adjective
maternal	0	Medium	Adjective
dashing	0	Low	Verb
implement	0	High	Verb
saxophone	0	Low	Noun
malevolent	0	Low	Adjective
insecure	0	Low	Noun
void	0	Low	Noun
rampant	0	Low	Adjective
lethal	0	Low	Adjective
terminal	0	Medium	Adjective
resented	0	Low	Verb
mosquito	0	Low	Noun
dehydrate	0	Low	Noun
mantel	0	Low	Noun
bestow	0	Low	Verb
murdered	0	High	Verb
punish	0	Low	Verb
toxic	0	Medium	Adjective

