YouTube in the EFL Classroom and Listening Comprehension

A Corpus-Based Study

Malin Sandsmark

Master's Thesis

Department of Foreign Languages

University of Bergen

2019

Acknowledgments

I would like to give a special thanks to my supervisor, Kimberly Marie Skjelde, for your professional guidance and valuable support. Thank you for sharing your own experiences so that I never had to feel alone during this process.

Thank you to my friends whom I have come to know these past five years. We did it! Thanks especially to Janne Helgesen, for not only being a good friend but for also being my biggest cheerleader throughout these five years. Thank you for your invaluable comments on this thesis.

Thank you to my family for always letting me choose my own paths and for supporting me endlessly.

Finally, I want to thank Severin for always saying the right things, for listening to my endless complaints, and for giving me the space to do my thing. I will forever be grateful for your patience during this project.

#lektorlove

<div align="right">

Malin Sandsmark

Bergen

May 2019

</div>

Abstract in Norwegian

Bruk av YouTube-videoer i klasserommet har blitt mer og mer vanlig siden opprettelsen av YouTube i 2005. Spesielt i engelskundervisningen kan slike videoer åpne et vindu inn til den verden som engelsktalende mennesker lever i.

Som lektorstudent har jeg fått et inntrykk av at lærere liker å bruke YouTube-videoer fordi de tror at elever vil forstå mer av videoer enn av en vanlig tekstbok, samt at videoer bidrar til variasjon i undervisningen. Videre har jeg også inntrykk av at lærere velger slike videoer ut ifra innhold – ikke etter hvor mye de forventer at elevene skal skjønne. Slike videoer krever mye av elevenes lytteferdigheter og de må prosessere alt som skjer på skjermen i løpet av kort tid. Uten å tenke over det kan lærere ende opp med å velge videoer som setter høye leksikalske krav til elevene fordi ordforrådet som kreves for å forstå videoene er høyere enn hva elevene har, og/eller fordi videoene inneholder alt for mange elementer på en og samme tid. Det er dessverre lite forskning på bruk av YouTube-videoer i engelskundervisningen. Denne oppgaven ønsker å sette søkelyset på dette feltet, her ved å undersøke de leksikalske kravene lærere plasserer på elevene sine når de viser slike videoer i klasserommet. Den ønsker også å undersøke hvordan ord blir presentert (som for eksempel skrevet tekst eller tale) og hvilke konsekvenser dette har for elevers lytteforståelse av videoene.

Oppgaven er en korpus-basert studie bestående av transkripsjoner fra 20 BBC News videoer fra YouTube. Det teoretiske rammeverket er basert på kognitive teorier som forklarer hvordan mennesker prosesserer lyd og bilder og hvilken påvirkning ulike elementer i en video kan ha på en elevs lytteforståelse.

Resultatene viser at et ordforråd på 3,000 ordfamilier gir kjennskap til 95% av innholdet, hvorav 5,000 ordfamilier gir 98% forståelse. Samtidig er tale den mest brukte måten å presentere ord på i videoene etterfulgt av skrevet tekst. Halvparten av videoene bruker både tale og tekst samtidig for å presentere de samme ordene. I lys av prinsipper for multimedia læring kan tale sammen med bilder øke forståelsen. Blir tale, tekst og bilde brukt samtidig kan dette forstyrre elevens oppmerksomhet og føre til redusert forståelse. Man kan også anta at ord som blir brukt i to moduser for å presentere den samme informasjonen er overflødig. Studien viser at lærere må være bevisste på hvilke krav de setter til elevene sine i møte med videoer. Dette fordi videoene kan være leksikalsk utfordrende, samtidig som måten ordene blir vist på enten kan være til hjelp eller til hindring av forståelsen.

Table of Contents

List of Abbreviations

EFL   English as a Foreign Language

ESL   English as a Second Language

L1   First Language

L2   Second Language

SLA   Second Language Acquisition

List of Tables and Figures

List of Appendices

1. Introduction

"We need technology in every classroom and in every student and teacher's hand, because it is the pen and paper of our time, and it is the lens through which we experience much of our world." - David Warlick, Education, Technology Educator[1].

1.1 Aims and Scope

Videos are multimodal texts consisting of sounds and visual images which can be presented using a range of media (Cross, 2011). With the arrival of video-sharing websites such as YouTube, English language teachers are provided with an abundance of contemporary visual and verbal illustrations of the cultures where the languages that they teach are spoken. Today, several types of videos such as documentaries, news segments, crash courses, and cartoons are readily available to use in the language learning classroom. Furthermore, videos may provide longed-for variation for both teachers and learners in the classroom. When videos are presented in the classroom, learners are required to listen. However, in a language learning educational context, a lot is known about reading comprehension while less is known about listening comprehension.

Videos exist in a variety of different styles with a variety of presentation modes, i.e. the way information is presented (audio, written text, pictures), so that when listening to videos, language learners must process multiple input modes simultaneously. Thus, using videos in the language learning classroom may aid learners when encountering a new language, as the visual element may provide them with contextual cues. This is further supported in empirical research, which has reported that videos support listening comprehension for (Baltova, 1999; Danan, 1992; Plass & Jones, 2005; Rubin, 1995; Siddell, 2011)

Furthermore, there might be reasons to believe that shorter videos, such as videos found on YouTube, will try to maximize the available time by presenting multiple modes of information, so that videos efficiently can present their message. However, cognitive theories, which claim that humans have limited capacity in their working memory to process information, suggest this is not necessarily so, as humans can only hold and process a few elements of information at a time (Mayer, 2014b; Mousavi, Low, & Sweller, 1995; Paivio, 1986; Sweller, Ayres, & Kalyuga, 2011). For example, Inceçay and Koçoglu (2017)

---

[1] http://2cents.onlearning.us/

examined whether or not different input delivery modes had an effect on listening comprehension of Turkish university students learning English. They found that when audio, video, and subtitles were combined, learners' comprehension were affected negatively. Thus, there seems to be a disagreement amongst scholars whether videos will facilitate listening comprehension or not. However, more research is needed in this area.

Additionally, videos depicting authentic language situations from the country of the native speakers are rarely made for second or foreign language learning and may be very difficult to comprehend. Through my own education and, later, as a teacher in training, I have observed that teachers seldom take the videos' language level into consideration when selecting videos as teaching materials. From what I have experienced, teachers search for the appropriate topic on video sharing sites, such as YouTube, and select a video based on its content and how it fits with the topic. Videos are popular to use in the classroom because the general assumption, in my experience, is that teachers believe videos are easy for learners to comprehend. Due to the visual elements videos provide, teachers might think that they provide learners with contextual support, which will aid their comprehension of the videos and fill in the gaps of missing vocabulary. However, little is known about the use of different elements (such as pictures, written text, animations, etc.) in videos such as these. The questions to ask are: when teachers use videos in the classroom, how can they know what demands they place on their learners are, and what are the possible impacts these videos can have on learners' listening comprehension?

Thus, the present study seeks to investigate listening comprehension in a multimedia context to better understand the relationship between lexical knowledge, video design, and L2 learners' listening comprehension of videos used in the classroom. The hope is to raise teacher awareness to the use of videos in the classroom, so that teachers will consider how lexical knowledge and input modes may affect L2 learners' listening comprehension. Thus, the current study is two-fold. Firstly, it will examine the lexical demands placed on L2 learners when videos found on YouTube are presented in the classroom. Secondly, it will explore the use of input modes in these videos to provide a better understanding of how input modes may help, or hinder, listening comprehension. The study will focus on videos for use in the English learning context of lower secondary education in Norway.

1.2 Centrality of News Videos as Teaching Materials

News videos are a valuable and commonly used resource in the classrooms for improving language learners' listening abilities, and a growing number of newspapers continue to offer

productions which can be utilized as material in the classroom (Gruba, 2006). BBC, for instance, has a long history of producing language materials even before the digital age, and now offers videos available online.

The British Broadcasting Corporation (BBC) is a British public service broadcaster and is the largest broadcasting corporation in the world. According to its website, BBC's mission is "to enrich people's lives with programmes that inform, educate and entertain" (BBC, 2019). BBC provides services such as television programs, radio programs, and online content. The broadcaster produces longer documentaries, shorter video clips, and also "learning"-content which is aimed at young learners. BBC, therefore, is very popular to use for educational purposes. Online content is available for anyone on the BBC website as well as on BBC's YouTube channels. BBC creates a variety of different videos related to various topics.

However, the primary purpose of sites such as BBC is journalism, as opposed to language pedagogy, and it is precisely the non-pedagogical focus that makes the videos popular as L2 teaching materials. For example, in the Norwegian educational context, news videos fit well into teaching the English subject curriculum as these videos expand on the communicative nature of the curriculum focus. The English subject curriculum emphasizes the need to use English for communication purposes, which also includes the ability to take cultural norms and conventions into consideration (The Norwegian Directorate for Education and Training, 2013). Produced by native speakers in the community of the target language, these videos provide authentic sociocultural and linguistic resources for the language learning classroom (Siddell, 2011). Thus, the authenticity news videos provide may contribute to a greater understanding of the English-speaking world.

The ARK&APP project[2] (Gilje et al., 2016) researched how teachers use educational materials and resources in four school subjects (English, Mathematics, Natural Science, and Social studies) at three levels in primary and secondary education in Norway. The results of the teacher survey suggested that the educational material most frequently used in the Norwegian classroom is generally the printed textbook. Digital educational resources are more frequently used in upper secondary school than in primary school, according to the survey. In primary school, 60% of teachers stated that they mainly use paper-based educational resources when teaching, while they make use of additional digital resources when they see fit. In the English subject, on the other hand, less than 50% of teachers report

---

[2] See more of the Ark& App project: https://www.uv.uio.no/iped/forskning/prosjekter/ark-app/

that they mainly use paper-based educational resources in teaching. The ARK&APP project illustrates how digital resources in the classroom are widely used by teachers, especially in the upper secondary levels. Thus, it can be assumed that news videos might be a part of Norwegian teachers' digital resources they use in their classroom for the English subject – in both lower and upper secondary education context. Nevertheless, there seems to be little research examining exactly what type of digital resources Norwegian English teachers actually make use of in the classroom.

1.3 Why Listening?

Listening is a critical second language (L2)[3] skill and has a fundamental role in the language acquisition process, but for many L2 learners, listening poses a grave challenge (Yang & Chang, 2014). There are many reasons for why L2 learners may struggle with comprehending spoken input. For instance, spoken language exists in time rather than space, and it must be understood simultaneously as it is uttered by the speaker (Lund, 1991). At the same time, the L2 learner must identify words in a continuous stream of speech, with no white spaces available for separating words, as there are when reading (Goh, 2000).

One key conceptual issue in defining listening as a skill involves the role of visual elements. When videos are shown in the classroom, they require the L2 learners to listen with their ears and see with their eyes. They listen to the audio recording and view the visual images presented in the video. Listening to a video may, therefore, be better described as video-based listening which Rubin (1995) defines as "an active process in which listeners select and interpret information which comes from auditory and visual cues in order to define what is going on and what the speakers are trying to express" (p. 7). Nevertheless, listening when communicating with another human being is also a process of listening with the ears and viewing with the eyes. The learner must interpret information from verbal and visual cues because the speaker will use non-verbal cues when communicating, such as gestures, facial expressions, and through the physical way words are coming from the mouth (Rost, 2016). For the purpose of this thesis, however, the term listening will follow Rubin's (1995) definition and, thus, includes both listening to verbal input and visual input.

Reasons for why teachers should use videos in their teaching can be argued for using Nation's (2017) four strands of a well-balanced language course. He suggests that when planning a language course, teachers should include an even balance of four strands:

---

[3] Second language (L2) will in this thesis refer to any language learned after the first. See Section 1.4.1.

meaning-focused input, meaning-focused output, language-focused learning and fluency development. Teachers should include an equal amount of activities from each of these strands because each language skill is different and will need to be attended to (Nation, 2017).

The meaning-focused input strand involves learning through listening and reading, which is the primary focus for the present study. Thus, this is the only strand that will be presented in this section. The main focus and interest for L2 learners should be on increasing their knowledge, understanding, and/or finding pleasure from what they listen to and read (Nation, 2017). Nation (2017) argues that the strand only exists if these specific conditions are present: learners are already familiar to what they are listening to or reading, the learners are interested in the input and motivated to understand it, only a small amount of the language features are unknown to the learners, the learners are able to acquire more information about the unknown language items through contextual cues and background knowledge, and learners are exposed to large amounts of input. Conclusively, it will not be meaning-focused input if these conditions are not present.

Listening to videos can provide L2 learners with valuable meaning-focused input, where the input is interesting, meaningful, and authentic while providing information about the culture and society for the speakers where the L2 is spoken. In Norway, the Purpose Section of the English Subject Curriculum provides many arguments which can be applied to the rationale for using videos in the language classroom:

> [i]n addition to language learning, the subject of English shall contribute to providing insight into the way people live and different cultures where English is the primary or the official language. … Learning about the English-speaking world and the increasing use of English in different international contexts will provide a good basis for understanding the world around us and how English developed into a world language. … Oral, written and digital texts, films, music and other cultural forms of expression can further inspire personal expressions and creativity. The subject shall help build up general language proficiency through listening, speaking, reading and writing, and provide the opportunity to acquire information and specialized knowledge through the English language (The Norwegian Directorate for Education and Training, 2013, p.1)

In contrast, certain characteristics of videos, especially news videos, may cause numerous challenges for L2 learners when attempting to comprehend them. For instance, the different

patterns of the discourse, unfamiliar vocabulary, speech rates, prosody, and other syntactic structures, the high concentration of factual content, or differences between aural and visual information may all cause difficulties for L2 learners (Cross, 2009). In addition, the presentation of several input modes may cause difficulties when processing the videos for the L2 learners (see Section 2.5). Hence, exactly how videos may or may not facilitate L2 learners listening comprehension remains unclear.

1.3.1 Why word knowledge?

The main idea that scholars agree on is that the greater the proportion of a text that is known, the better understanding will be. Thus, the more words learners know, the better they will be able to understand when reading or listening in the L2. For teachers, L2 learners' word knowledge can be a great indication for how much they can expect their learners to understand when they present videos in their classroom. Hence, vocabulary knowledge is the most important area of knowledge required for successful listening comprehension (Milton, 2009), which is why this will be the focal point for the current study.

Although there are several studies examining how much vocabulary knowledge is required to comprehend certain discourse, majority of the research has been conducted on reading. Therefore, there seems to be a shortage of research in the area of L2 listening comprehension. While Webb and Rodgers (2009a, 2009b) assessed how much vocabulary knowledge is needed to comprehend movies and television programs, and Nation (2006) assessed the lexical demands of the movie *Shrek*, to the best of my knowledge, no research has been conducted on the vocabulary size needed to comprehend videos found on YouTube. Thus, it is my hope that the study's investigation will contribute to fill this void.

1.4 Terminology

1.4.1 The situation of English in Norway

Norwegians are not considered speakers of English as a second language (ESL) under traditional descriptions because "they are neither speakers of new Englishes in postcolonial countries nor immigrants to a native- English-speaking country, and English does not have official language status in Norway" (Rindal & Piercy, 2013, p. 212). However, with the increase of English as a global language, Norwegians are exposed to English in their daily lives outside of the school situation. Consequently, the traditional EFL model seems to be in decline (Graddol, 2006) because the status of English in Norway is no longer that of a foreign

language, but not entirely a second language either, and, hence, seems to be caught between English language paradigms (Rindal & Piercy, 2013). Thus, it seems more appropriate to use the term second language, which in this context will be defined as any language acquired after the first (Herschensohn & Young-Scholten, 2013). For the present study, L2 will refer to this definition, where L2 learners will refer to learners who learn any language after the first. Thus, Norwegians will fall under this term, as they generally learn English after Norwegian and will in this thesis be included when using the L2 term. However, the traditional EFL term will be used when referring to the Norwegian English-learning classroom as the situation in Norwegian schools is still in line with the traditional EFL model.

1.4.2 Multimedia, modality, and modes

Multimedia refers to the use of more than one presentation medium, such as words and pictures (Mayer & Sims, 1994). Mayer (2014c) provides a broad definition of multimedia as presenting both words (such as printed text or spoken text) and pictures (such as illustrations, photos, animation, or video). Furthermore, he adds learning to the concept to form multimedia learning. Thus, multimedia learning refers to the learner's construction of knowledge using words and pictures in his building of mental representations. Additionally, multimedia instruction involves presenting words and pictures to promote learning. In other words, it means to design a multimedia presentation that aids learners in creating mental representations. The terms modality and modes will, in the present study, follow Mayer's (1997, 2001) definitions, where modalities refer to the verbal and visual working memory channels where input enters, and modes will refer to the way this information is presented, such as through speech, music, written text, images, animations, and so on. Single-mode will refer to the use of one presentation mode in one modality (e.g. spoken text in the verbal channel), while dual-mode refers to the use of two modes in one modality (e.g. spoken text and on-screen text in the verbal channel).

1.5 Outline of the Thesis

The thesis is structured into five chapters. The first chapter has introduced the aims and scopes of the present study and placed it in a research gap. The second chapter will present the theoretical framework and relevant research that will guide the rest of the study. The third chapter will outline the methods and analyses used to conduct the research, while the fourth chapter presents the findings and the discussion of the findings. Lastly, the final chapter will

conclude the present study, provide limitations of the thesis, provide suggestions for further studies, and summarize the pedagogical implications.

As stated, to the best of my knowledge, there are no other studies that have examined the lexical demands of BBC-videos or other videos found on YouTube, nor have I come across any studies investigating the use of different multimedia presentation modes in such videos. I intend to fill this research gap with the present study, by asking the following research questions:

1. What are the lexical demands placed on Norwegian secondary L2 learners when watching BBC News videos available on YouTube?
2. How are different input modes presented in these videos?
   2a. How can the use of input modes in these videos impact listening comprehension?

2. Theoretical Background

This chapter intends to give a brief overview of theories relevant to the present study, as well as outlining the previous research. Finally, both the theoretical aspects and previous will be related to the aim and research questions for this thesis.

2.1 Theoretical Framework

Second language acquisition (SLA) is a complex process, and there is no single way in which a learner acquires knowledge of a second language (Ellis, 1986). However, VanPatten and Williams (2007) suggest that, between the advocates of different SLA theories, there is an agreement that (1) exposure to input is vital; (2) a great proportion of SLA happens without awareness; (3) the stages of acquisition are predictable; (4) frequency is essential; (5) the first language affects SLA; (6) speaking affects SLA; and (7) linguistic knowledge affects SLA. In the present study, I have chosen to focus on the processes involved when listening to input in an SLA context. Specifically, the exposure to different input modes in videos and how linguistic knowledge may affect listening comprehension of these videos. The following sections will outline the processes involved when listening to input and discuss the relationship between linguistic knowledge and listening comprehension. Furthermore, the theoretical framework guiding the present study, which is based on the cognitive load theory (Sweller, 1988) and multimedia learning theory (Mayer, 2001) and their relevant principles, will be presented. For the purpose of the current study, the principles are limited to the multimedia principle, the split-attention principle, and the redundancy principle. These principles will be explained in detail in Sections 2.5.3.1, 2.5.3.2, and 2.5.3.3.

2.2 Input

This section aims to define the term input to provide a deeper understanding of why input is important in the context of listening comprehension, and why it is relevant to this study.

Input can be defined as the language that learners are exposed to when, for instance, they are listening or reading (Gass & Mackey, 2014). Most importantly, input contains a message for learners to attend and respond to (Barcroft & Wong, 2013).

Input be presented in a range of modes, for instance, as written text, speech, images, and animation (see section 1.4.2). In a multimedia discourse, information is presented through a combination of visual and verbal modes, such as words and pictures. It is the multimedia

presentation of input that is of concern in the present study, as a video is one form of multimedia presentation combining visual and verbal input modes.

Since the late seventies and early eighties, the role of input in SLA has been the subject of much debate, especially with Krashen and his Input Hypothesis. He argued that comprehensible input is necessary for SLA to occur, where he defined comprehensible input as input that is slightly above learners' present levels of competence (Krashen, 1985). Thus, input which is too simple or too complex will not be useful for L2 learners when learning a language. Although his hypothesis is criticized for being untestable and vague (Mitchell, Myles, & Marsden, 2013), his view emphasized the importance of comprehensible input and has impacted how researchers in the field of SLA view the role of input (Barcroft & Wong, 2013). The important role of input in a language course is also expressed by Nation's (2007) four strands, where exposure to meaning-focused input make up one strand of a balanced language course (see section 1.3.1)

In the present study, the input processes that are involved when listening to input will be the main focus. The study does not assess learners' comprehension of the videos but investigates more closely the lexical demands placed on L2 learners and the mental processes these learners engage in when listening, in order to understand how exposure to videos in the classroom impacts L2 learners' listening comprehension. Furthermore, the present study concerns comprehension of input, and so does not involve production of output.

2.3 Listening

Listening is an intricate cognitive process that permits humans to comprehend spoken language. When presenting videos in the EFL classroom, learners are generally exposed to spoken language that require them to use their listening skills. While this is not always the case, as there are videos where spoken text is not present, listening will be the focal point for the rest of this thesis. Consequently, this section will try to define what is meant by listening in order to better understand what processes are involved when L2 learners are exposed to video in the classroom.

Listening that involves input from a linguistic source is seen as a fundamental aspect of listening to language (Rost, 2016). It involves "receptive, constructive, and interpretive features of cognition, which are used when listening in both the first language (L1) and second language (L2)" (Rost, 2005, p. 503). In L1 acquisition, listening skills and cognition are regarded as developing conjointly. As a consequence, listening is seldom given direct attention as a specific skill in L1 education. In an L2 context, however, direct attention is

viewed as essential for learning the language, because, in most cases, the learner is learning the second language after cognition and conventions in the L1 have been established (Rost, 2005).

When listening, learners engage in top-down and bottom-up processes which are simultaneous processes taking place at different levels of a learner's cognitive organization (Rost, 2005). When engaged in bottom-up processing, the learner recognizes and decodes the units of input language in order to comprehend the sentence or utterance. When engaged in top-down processing the learners use their prior knowledge and world experiences to build a conceptual framework for comprehension (Loewen & Reinders, 2011). Top-down processing can help bridge the gap in comprehension without depending heavily on linguistic features, as learners do when involved in bottom-up processing (Vandergrift & Goh, 2009). Nevertheless, scholars have found that, while readers rely more on information found in the linguistics features of the text (Reves & Levine, 1988), listeners rely more on top-down processing, e.g. using previous world knowledge to identify main ideas of a text (Lund, 1991).

When listening, learners engage in three basic processing phases that are simultaneous and parallel: decoding, comprehension, and interpretation (Rost, 2005). That is, to be able to listen, L2 learners need to decode the input, comprehend the input, and interpret the input. Each of these phases contributes to finding what is relevant to the learner in the input and will be presented in the following three sections.

2.3.1 Decoding

The first phase involved when listening is decoding. Decoding includes attention, speech perception, word recognition, and grammatical parsing (Rost, 2016). For processing of information to occur, learners must first pay attention to the input. Thus, attention is a process of directing the selection of input to make the decoding process more efficient. L2 learners cannot process all the input they are exposed to (Gass & Mackey, 2014). The working memory is quite limited both temporarily (how long one can process input) and semantically (how many separate items one can process at a time). Therefore, L2 learners need mechanisms, such as attention, to help them organize the input that they need whilst disregarding input they do not need (Rost, 2016).

The second phase of decoding the input is perception. Perception, in this context, is viewed as "the ability to recognize the auditory and visual components of language" (Loewen & Reinders, 2011, p. 133). In a listening context, Rost (2016) argues that the learner needs to identify the physical ways the speaker is making the input (how the mouth is working to make

the sounds), the aural qualities of the input (such as stress, pitch, intonation), and what the speaker is trying to articulate (the information coming out of the mouth). These are three complementary ways of perceiving speech. Speech perception is aided by the visual cues from the speaker, such as gestures and movement when articulating. When auditory cues are absent, however, there seems to be consistently more problems with comprehension (Massaro, 1998). Ninety percent of all content words in English have stress on the first syllable in English while non-content words are usually un-stressed (Cutler & Carter, 1987). Thus, a skilled listener can use stress to identify and organize word recognition around the stressed syllables as a way to segment words (Rost, 2016)

The third phase of decoding input is recognizing words. Lexical segmentation is the processes of recognizing words in the stream of speech, and when listening, lexical segmentation involves two synchronous tasks of the listener: first, recognizing words and lexical phrases and, second, activating knowledge associated with those words and phrases (Rost, 2016). Nevertheless, the background knowledge activation needed for comprehension of speech is made possible by word-recognition. In uninterrupted speech, however, there are no aural "white spaces" as there are in text, so the learner does not have helpful indications for recognizing word boundaries. It is often word recognition that becomes the most difficult process of decoding input for L2 listeners (Rost, 2016). Even if they know a word, L2 listeners may not always be able to recognize them in continuous speech (Vandergrift & Goh, 2009).

Word knowledge (see Section 2.4) involves numerous aspects and constantly expands (Rost, 2016). When a learner knows a word, it includes not only to be able to recognize the spoken and written form of the word and its grammatical functions, but also to be able to recognize the word's collocations, its frequency, denotations, connotations, association, concepts and referents (Schmitt, 2000). In principle, listening is dependent on the size of a learner's vocabulary and his ability to recognize words (Rost, 2016). Consequently, word recognition and vocabulary knowledge play an important part in second language listening comprehension, and thus, play an important role in the present study. Vocabulary knowledge and vocabulary size will be further discussed in Section 2.4.

Difficulties with word recognition were found in a study conducted by Rost and Ross (1991), where L2 learners reported that they often were able to identify a word but had problems with not knowing it or not able to recognize the word boundaries. They also found that

[a]dvanced learners apparently formulate discourse-level inferential questions with relative ease because they can allot more attention to the overall story structure. Beginning learners, on the other hand, lacking a critical mass of word knowledge, are seemingly forced to allot most of their attention to specific word meanings and parsing the input into basic constituent structure (p. 262).

Thus, the importance of word knowledge and attention in the decoding process is further supported. Goh (2000) identifies real-time word segmentation as the main difficulty for L2 learners' listening comprehension. The spoken discourse has, in addition to speed, linguistic features that learners often find difficult to process such as elision, reduction, and assimilation. Also, cognates that are alike in print but have phonetic differences may be difficult for the listener to decode correctly (Brown, 2006). While readers can look for contextual clues within the text, listeners cannot attend to a single word without losing the message that follows, which may cause difficulty in decoding the input (Rost, 2016).

Decoding input is an important phase for comprehension to occur. That is, learners cannot comprehend discourse they have not decoded. As the empirical evidence puts forward, attention and word knowledge seem to be the main obstacles for successful decoding for L2 learners when listening. This finding stresses the need for teachers to have knowledge of the processes involved when L2 learners listen to videos, in order to understand factors that may impact their listening comprehension.

2.3.2 Comprehension

The second phase involved when listening is comprehension. The present study has placed an emphasis on L2 learners' listening comprehension when watching videos, and, thus, a more in-depth description of the processes involved provides a better understanding of how comprehension may or may not occur.

Listening comprehension is a complex term because the term "comprehension" has not been satisfactorily defined in L2 research (Baggett, 1989). Thus, this study will not go into details about degrees or levels of comprehension, but follow Rost (2005) who understands listening comprehension as a process where the learner "construct[s] concepts in memory for the lexical items uttered and propositions stated or implied by the speaker" (p. 512). In L2 learning, a considerable number of concepts has been acquired through the L1. Hence, unlike L1 speakers, L2 learners must develop new forms and create new links between the new forms and their current concepts (Tschirner, 2011). Furthermore, according to

VanPatten (2014), acquisition cannot occur without comprehension. That is, if L2 learners cannot comprehend the language that they are exposed to, then that language is not useful for them in the process of acquiring the L2. Furthermore, unlike L1 native speakers, L2 learners must actively attempt to develop their ability to comprehend (VanPatten, 2014). Consequently, comprehension is a process of constructing real world language references to concepts and representations in L2 learners' personal memory (Rost, 2005). It involves four overlapping sub-processes: identifying salient information, activating appropriate schemata, inferencing, and updating representations.

Listening comprehension involves an ongoing cycle where each utterance by a speaker contains both new and given information where a listener attends to the information given by the speaker and identifies salient information (Rost, 2016). The cues of what is new and what is given information can be found in the prosody of speech. However, it is important to note that mutual comprehension is never guaranteed during this process, as there can be discrepancies between the speaker's concepts and those of the learner (Rost, 2005).

Another central component in comprehension is the activation of appropriate schemata (Rost, 2016). Schemata are mental structures that learners use to organize knowledge of concepts. These structures allow learners to recognize the whole picture of an event or object from limited information provided by the input, which, when receiving information, activates the appropriate scheme. According to Rost (2005), "[u]nderstanding what a speaker says depends to a large degree upon shared concepts and shared ways of reacting to the world, or at least the imagination of shared concepts" (p. 513). In L2 listening, however, learners come in contact with assumptions and expectations that are in disagreement of their own. It is, therefore, important for L2 learners to activate the appropriate schemata when listening so that comprehension problems do not occur (Rost, 2016).

The process of inferencing is also an essential element which has an impact on listening comprehension. Inferencing is the process where a learner assumes a conclusion to something based on related propositions, which he often does based on incomplete evidence (Rost, 2005). When meeting unknown words, learners might use strategic inferencing to compensate for this by making use of the visual and aural cues (Vandergrift, 2007). Theoretically then, comprehension occurs when listeners are able to infer meaning from speech utterances based on their linguistic background and contextual knowledge (Buck, 2001).

Finally, comprehension involves the construction and updating of concepts in memory for the lexical items received when listening. Rost (2016) proposes two dimensions

of memory usage; short-term and long-term memory. Short-term memory is more recently known as working memory (Mayer, 2005), which is the term that will be used in the present study. Working memory contains knowledge that is activated at a particular time. Long-term memory, however, contains the sum of all knowledge a learner possesses, which is inactive knowledge, but can be activated by bringing it to the working memory. Comprehension, thus, occurs in the working memory, and a learner's ability to deal with input instantly in the working memory is a necessary component of listening competence in the L2 (Rost, 2016). The role of memory will be further discussed in section 2.5.1 with the cognitive theory of multimedia learning and cognitive load theory.

2.3.3 Interpretation

Interpretation includes comparison of meanings with prior expectations, activation of participation frames, and evaluation of discourse meanings. This is the stage of listening where the listener positions himself to the speaker's meaning by taking on the perspective of the speaker, as well as assessing how it has relevance for him (Rost, 2005). Thus, interpreting spoken language includes the learners situating himself "in terms of the topic, the setting, the event, the speaker and purpose for listening and relationship to the topic, setting event and speaker" (Rost, 2005, p. 519). Consequently, because of differences in social knowledge, different listeners understand different things from the same input. Listening is, thus, a co-construction of meaning (Rost, 2016). For the purpose of the current study, the processes involved when decoding and comprehending input when listening will be the focal point for the discussion of the findings. Although Rost (2005) argues that these three processes are intertwined, interpretation is involved after decoding and comprehension of an element have occurred. Consequently, the interpretation stage has only been briefly discussed in this section, and a discussion of interpretation and listening to videos will not be within the scope of the present thesis

2.3.4 Summary of processes involved when listening

In essence, listening comprehension is a complex, inferential process (Rost, 2002). The learner decodes, comprehends, and interprets spoken messages simultaneously as the messages are presented, and the listener engages in top-down and bottom-up processes to build mental images in memory of what has been comprehended (Vandergrift, 2006). Listeners, unlike readers, need to deal with spoken language. Hence, they must attend to

several factors that can complicate the process of listening comprehension which can be cognitively demanding (Buck, 2001).

The processes involved when listening are provided in details in the present study because the study aims to investigate how L2 learners' listening comprehension may be impacted by the videos shown in the classroom. Teachers need to be aware of the processes involved when L2 listeners are presented with a listening task in the classroom to better understand how the teaching material will, or will not, be successfully comprehended.

2.4 Word Knowledge

As discussed in the previous section, word knowledge is an essential part of the process involved when listening to input. This section will present Corpus Linguistics, which has been used as a framework for assembling the corpus in the current study. The section will further review what it entails to know a word and how many words it can be expected for L2 learners to know in order to understand videos found on YouTube which are shown in the classroom.

2.4.1 Corpus Linguistics

With developments in computer technology, large collections of language data can be stored electronically (Tognini-Bonelli, 2010). Corpus Linguistics provides insights of language usage that were not available before the invention of computers. However, "[C]orpus [L]inguistics did not begin with the development of computers …" (Kennedy, 1998, p. 2), as collecting texts for analysis has been done manually for years.

In contrast to other disciplines in linguistics, Corpus Linguistics does not study any particular aspect of language (McEnery & Hardie, 2012). There is no exact definition related to the use of Corpus Linguistics within the field of linguistics. This is due to the fact that researchers do not agree on its scientific status (Taylor, 2008). In other words, if Corpus Linguistics should be regarded as theory or as methodology. Tognini-Bonelli (2001) claims that Corpus Linguistics is somewhere in between, and that it

> has been argued that [C]orpus [L]inguistics is not really a domain of research but only a methodological basis for studying language …[while] [m]any linguists working with a corpus, however, tend to agree that [C]orpus [L]inguistics goes well beyond this purely methodological role (p. 1).

Thus, there is a difference between corpus-driven and corpus-based approaches to Corpus Linguistics (Tognini-Bonelli, 2001). Corpus-driven research begin with the data from the corpus and derive the theory on the basis of these findings. Thus, the data becomes the theory.

Corpus-based research, on the other hand, set out to test a hypothesis with the use of corpora, basing the study on already existing theories (McEnery & Hardie, 2012). In the current thesis, a corpus-based approach to language research has been applied. This will be further described in Section 3.2.

A corpus can be defined as a "collection of texts assumed to be representative of a given language put together so that it can be used for linguistic analysis" (Tognini-Bonelli, 2001, p. 2). A corpus can be used to analyze natural and authentic language, and to investigate patterns of language use within a range of texts. Corpus Linguistics is used as an approach to corpus work by being an empirical approach to the description of language use, operating within the framework of a "contextual and functional theory of meaning", and making use of the new technologies (Tognini-Bonelli, 2001, p. 2). Given that a corpus is a collection of texts, the aim of Corpus Linguistics can be seen as "the analysis and the description of language use, as realised in text(s)" (Tognini-Bonelli, 2010, p. 19).

For the present study, my focus will be on the use of Corpus Linguistics as a method to gain numeric data related to the use of vocabulary in videos found on YouTube, in other words, applying Corpus Linguistics as a research tool for building my corpus. Thus, as previously mentioned, the present study takes on a corpus-based approach to investigate language in use. A further discussion of the corpus that was assembled for the purpose of this study is provided in section 3.1.5.

2.4.2 Frequency

In vocabulary learning and acquisition, frequency is the most important feature, affecting almost all aspects of lexical processing and acquisition. Corpus data is said to be the most suitable source of finding frequency information (Schmitt, 2010, p. 13). The main idea behind the use of corpora in language research is that the words in a language are not created equal. That is, some words are more essential to learn than others (Nation, 2001). As such, by looking at a word's frequency one can determine the importance of a word (Nation, 2001). This is especially useful for teachers as they can concentrate on teaching a smaller number of words to L2 learners, and still make sure that the learners will be able to understand and be understood.

Moreover, frequency indicates how often words appear in a particular discourse and can be a useful measurement when assessing lexical demands, because the more frequent a word is, the more it will appear in any discourse, while less frequent words will appear less often (Nation, 2006). This can be illustrated through Zipf's law (1949), which indicates that

"the rank order of an item in a frequency-ranked list multiplied by its frequency results in a constant figure" (Nation & Webb, 2011, p. 131). Thus, Zipf's law demonstrates that a minor number of high-frequency words cover a large percentage of the discourse, while a large number of low-frequency words cover only a small percentage of the discourse (Nation, 2013). Conclusively, when researchers claim that having knowledge of high frequency vocabulary is important, they mean that these words have a higher chance of occurring in different use of language (Nation, 2001). Furthermore, because a small number of words cover a large percentage of a text, learners can gain successful comprehension even with a smaller vocabulary size, as long as the learner knows the right words (in this case, high-frequency words).

In order to establish that some words are more important to learn than others, several scholars have tried to establish the frequency of words in different discourse through the use of word-frequency lists (Coxhead, 2000; West, 1953). In these lists, words are ranked according to how often they occur in different types of discourse. Word-frequency lists are usually created using Corpus Linguistics, where samples of texts gathered from different corpora are analyzed for how often they occur. The software ranks words according to how frequently they occur in the corpus. For example, Nation (2017) created the BNC/COCA word family list based on the BNC and COCA corpus, which has been used in a variety of research. Additionally, in comparison to other lists, Dang and Webb (2016) found that the BNC/COCA lists performed well on both spoken and written texts. A more detailed description of the word-frequency lists is provided in section 3.1.2.

Nation (2006) suggests a division between high- and low-frequency words, where the 2,000 most frequent word are classified as high-frequency words, and less frequent words are classified as low-frequency words. However, Schmitt and Schmitt (2014) suggest a division between high-frequency, mid-frequency, or low-frequency, where words belonging to the first 1,000-3,000 word families are classified as high-frequency words, mid-frequency words are words between 4,000 and 9,000 word families, and low-frequency words are words from the 10,000 level and beyond. In the present study, the division proposed by Schmitt and Schmitt (2014) will be used.

Indeed, the distribution of high-frequency, mid-frequency, and low-frequency words, as illustrated by Zipf's law, reveals that it should be of great interest for teachers and learners of another language to learn the high-frequency words before the mid- and low-frequency words. If L2 learners do not know the high-frequency words, then they will encounter difficulties in all uses of the language and it will weaken their ability to communicate well

with others (Webb & Nation, 2017). Empirical research shows that learners generally learn high frequency vocabulary before they learn low frequency vocabulary (Read, 1988). This is especially important for the purpose of this study, because it illustrates how L2 teachers must make important decisions of what vocabulary they should make sure their learners will need to know.

Conclusively, teachers will need to make sure that their learners have sufficient knowledge of the right words. That is, L2 learners should know most of the high frequency words before moving on to learn mid-frequency words. If learners lack knowledge of the right words, it will weaken their comprehension of the input presented to them. In order to know what words will be most important for L2 learners to have knowledge of in order understand different discourse, teachers will need to have an idea of how large a vocabulary size their learners will need to acquire. Therefore, the following sections will discuss word knowledge and vocabulary size. Furthermore, the last section will discuss the relationship between vocabulary size and listening comprehension.

2.4.2.1 What is a word?

In order to count words according to frequency, one must decide what should count as a word. The English language contains an immense number of words and several scholars have tried to estimate a specific number. However, the number is continually growing and can only be an estimation (Schmitt, 2000). Words contain a base, root, or stem – the simples form of that word, where affixes are added to this base. Inflections refer to cases where the affixes are added for grammatical purposes, whereas derivates refer to cases where the affix changes the word class of a base (Schmitt, 2000). A set of word forms (the base word, all of its inflections, and its similar derivates) composes a word family. However, one must be cautious when deciding what affixes to include in a word family, as learners typically develop more knowledge of prefixes and suffixes over time. Bauer and Nation (1993) proposed a graded set of seven levels arranging the inflectional affixes and the most beneficial derivational affixes of a word base. The inclusion of an associated form of a word within a word family is centered around criteria such as "frequency, regularity, productivity, and predictability" (Bauer & Nation, 1993, p. 253). Lemmas, on the other hand, consist only of the base word, its inflections, and its reduced forms (Nation, 2013). In the current thesis, word families will be used as the unit of counting as Schmitt (2000) points to extensive evidence that the human mind groups members of the word family together that would be perceived as the same words. Furthermore, using word families as the unit of counting means that one assumes that

if the learner knows the headword, the learner will also know all of the family members of the word. Scholars state that this is true for the receptive knowledge of word families (which will be discussed below), because knowing one member of a word family may likely help learners recognize other members (Bauer & Nation, 1993). It is also suggested that when reading and listening, learners who know one of the members of a word family will probably know other family members by using their knowledge of common and regular word-building devices in English (Nation, 2006).

2.4.3 Receptive or productive vocabulary knowledge?

Listening and reading are seen as receptive skills, where learners receive language input (Nation, 2013). Thus, scholars agree that a distinction of word knowledge should be made between productive and receptive word knowledge. Productive knowledge involves producing language by speaking and writing, while receptive knowledge involves receiving language from input through listening or reading. Schmitt (2000) defines receptive knowledge as "being able to understand a word" (p.4), while productive knowledge of a word is seen as being able to produce a word on one's own accord (Schmitt, 2000). However, Nation (2013) argues that the distinction between the terms should not be placed on either end of a dichotomy, but rather in a continuum, as there are productive features in receptive knowledge when learners "produce" meaning when listening. Thus, receptive vocabulary knowledge, as proposed by Nation (2013), will in the present study be referred to as the ability to identify the form of a word when listening or reading and retrieving its meaning. More specifically, having receptive knowledge of a word involves being able to recognize the written and spoken word form, recognizing the word's meaning and concept in the context it occurs, recognize the word-parts making up the whole word, and knowing its related words (Nation, 2013).

Conclusively, knowing a word when watching a video concerns having receptive knowledge of a word by being able to make form-meaning connections. This distinction is central in the current study because when watching videos, L2 learners will need to have receptive knowledge to recognize and understand words. Furthermore, it is generally accepted that learners' receptive knowledge is greater than productive knowledge of words (Webb, 2008). This has implications for the lexical demands placed on L2 learners when teachers present videos, as learners will, most likely, have a larger receptive vocabulary size than productive. Hence, watching videos may be less demanding than situations which require

productive vocabulary, such as speaking, as it has been suggested that recognizing and understanding words are seen as less demanding than producing words (Schmitt, 2000).

2.4.4 Vocabulary size

For teachers to know how lexically demanding the videos they use in the EFL classroom are, it is important for them to have an idea of how large a vocabulary they can assume that their learners will need to have. Thus, to better understand how large a vocabulary one can expect L2 learners to need, it will be beneficial to have an idea of how many words native speakers know. This way, teachers and learners of an L2 are provided with a reference when determining an appropriate vocabulary goal.

Scholars have estimated that English native-speaking University students have a vocabulary size between 15,000 to 20,000 word families (Goulden, Nation, & Read, 1990). Nation and Waring (1997) established that

> [t]he best conservative rule of thumb that we have is that up to a vocabulary size of around 20,000 word families, we should expect that native speakers will add roughly 1,000 word families a year to their vocabulary size. That means that a five-year-old beginning school will have a vocabulary of around 4,000 to 5,000 word families. A university graduate will have a vocabulary of around 20,000 word families (p. 7-8).

However, L2 learners will probably never reach a native-sized vocabulary during their compulsory education, as the number of word families proposed by Nation and Waring (1997) is very high. Nevertheless, studies looking at native speakers' vocabulary growth weigh each word as having equal value, while frequency studies show that this is not necessarily true for the L2 learner (Nation, 2013). One cannot properly use a language to communicate if the vocabulary is insufficient. Consequently, researchers are concerned with how much vocabulary is necessary to enable communication (Schmitt, 2010). This is also important for teachers and learners of an L2, as they will need to decide which words should be prioritized and how many words should be an acceptable learning goal. The vocabulary goal should at minimum enable language for communication purposes. Several scholars have suggested such a vocabulary goal. For example, Nation (2006) suggests that knowledge of the most frequent 2,000 word families is essential for any language use, and, thus, for L2 learners knowledge of 2,000 word families should be the minimum vocabulary goal. This is supported by Adolphs and Schmitt (2003) who suggest that a vocabulary size of 2,000 word families

would suffice to *begin* engaging in everyday communication. However, how large a vocabulary one needs, depends of what type of discourse one sets out to understand (Webb & Nation, 2017). For instance, Nation (2006) suggests a vocabulary size of the most frequent 8-9,00 word families for understanding written texts, such as newspapers and novels, while Tegge (2017) found that a vocabulary size of 3,000 word families was necessary to understand lyrics of chart songs.

Nevertheless, there seems to be limited research measuring the vocabulary size of Norwegian L2 learners in both lower and upper secondary education. Studies from other Nordic countries, such as Iceland and Denmark, which assessed the vocabulary size of L2 learners may provide an idea of how the situation is in Norway. In a study testing the receptive vocabulary size of Icelandic tenth graders, Fleckenstein (2015) found that 59% of the participants' receptive vocabulary size consisted of 7,100-8,000 word families, while alarmingly 41% of the learners tested did not have the vocabulary size necessary to understand the authentic spoken and written language used in the classroom. Furthermore, Stæhr (2008) assessed the receptive vocabulary size of Danish L2 learners leaving the lower secondary level and found that around 77% of the learners had not acquired the most frequent 2000 word families in English – despite the fact that they had been exposed to seven years of English teaching. Additionally, Henriksen and Danelund (2015) found that only 48% of Danish L2 learners in Grade 13 had acquired knowledge of the 2,000 most frequent word families. The findings from these studies brings forward an impression that L2 learners might not have a large enough receptive vocabulary size to understand the material presented to them in the EFL classroom. However, more research is needed to claim that this is in fact the case. As this study's focal point is L2 learners in the lower secondary school in Norway, the results from the Danish and Icelandic studies suggest that not all lower secondary L2 learners will have the minimum requirement of vocabulary knowledge as proposed by Nation (2006). However, these figures cannot be directly applied to the situation in Norway, as there is not enough research to support this claim. As previously stated, the hope of this study is to gain more knowledge of how large a vocabulary size L2 learners *should* have in order to comprehend videos found on YouTube used as teaching materials.

Conclusively, the research illustrated above show how comprehension and vocabulary size are closely connected: learners with a larger vocabulary size are able to comprehend more of the input, where more comprehension of the input allows a learner to process more of it, which will lead to a development in vocabulary knowledge (Nation, 2013). However, this

is just a simplified way of seeing it. The processes involved when listening are more complex than this, as illustrated in Section 2.3.

2.4.4.1 Listening and vocabulary knowledge

Several scholars have tried to assess how large a vocabulary L2 learners will need to have in order to comprehend different discourse, and empirical studies have broadly documented the robust relationship between vocabulary size and reading. However, less research has been conducted on how vocabulary size impacts listening comprehension. This section will outline empirical research investigating size of vocabulary and listening comprehension, in order to gain an understanding of how these two are connected.

Estimates of vocabulary size are generally derived from studies of reading comprehension which have used coverage as a measurement to assess comprehension of different discourses (Adolphs & Schmitt, 2003; Nation & Waring, 1997; van Zeeland & Schmitt, 2012). Coverage will in this present study refer to the "percentage of words that a [learner] understands [in a discourse]" (Laufer & Ravenhorst-Kalovski, 2010, p. 16). Targets for coverage are usually set at 95% and 98% coverage, as extensive research on reading comprehension suggests these figures to be the "minimal" and "optimal" levels of coverage when reading (Hu & Nation, 2000; Nation, 2006; Schmitt, Jiang, & Grabe, 2011).

Thus, there seems to be a disagreement on whether the coverage percentages set for reading comprehension truly are representative for listening comprehension, considering the different characteristics between spoken and written discourse (van Zeeland & Schmitt, 2012). Nation (2006) investigated how much vocabulary was needed for reading and listening by examining the vocabulary size needed to understand different written texts, scripted spoken discourse, and unscripted spoken discourse. He used the coverage figures of 95% and 98% and found that, for scripted spoken English, learners needed to know 4,000 word families, plus proper nouns, to reach 96.74% coverage and 7,000 word families to reach 98.08% coverage. With unscripted spoken English, one needed 3,000 word families, plus proper nouns, to reach 95% coverage, and 6,000-7,000 word families to reach 98% coverage, which suggests that unscripted spoken English may be slightly less demanding than scripted spoken English. Compared to written English, however, research suggests that spoken language makes use of slightly more high-frequency words than written language, although one might need to consider a higher coverage than 98% due to the momentary nature of spoken language (Nation, 2006).

Hu and Nation (2000) count proper nouns as known words that does not need to be learned before reading or listening to the discourse, which is why Nation (2006) and other researchers have regarded proper nouns as known in their calculations of the vocabulary size needed to understand certain discourses. It is important to mention, however, that for L2 learners, proper nouns are usually different from that of their L1 language. Thus, L2 learners must, in addition to other types of words, recognize proper nouns as proper nouns, whereas L1 speakers not necessarily need to attend to these words. Even more importantly, for the purpose of the current study, Nation (2013) suggests that the language in newspapers and news stories has a larger proportion of proper names than novels (5-6% in newspapers, whereas 1-2% in novels) because news texts generally include interviews or comments from different people, which they often will need to present by including their names. This is important for teachers to be aware of when presenting news videos in the classroom, as L2 learners will need to recognize a higher proportion of proper nouns than when reading a novel. Consequently, teachers need to make sure that their learners know how to recognize proper nouns.

Furthermore, van Zeeland and Schmitt (2012) argue that coverage figures of 95% and 98%, suggested from reading comprehension research, cannot be directly applied to listening comprehension. In their study, they discussed whether the role of vocabulary knowledge would be smaller or larger with L2 listening than L2 reading, due to the differences between processing spoken and written text. L2 listeners might be affected by the characteristics of spoken language being of momentary character, and focus less on the unknown vocabulary, whereas the permanent feature of written text offers the reader the chance to refer back to unknown vocabulary to aid them when decoding the message (Reves & Levine, 1988). When considered individually, these factors imply that listening requires greater vocabulary knowledge than reading.

Nevertheless, researchers have found that readers rely more on linguistic information from the text than listeners (Reves & Levine, 1988). There are numerous characteristics of spoken language that may compensate for missing word knowledge. As previously mentioned, spoken input affords extra-linguistic support for comprehension with prosody and non-verbal information such as gesticulations, facial expressions, and lip movements (see Section 2.3.1). Spoken language also tends to be lexically less dense than written text in that it contains many fillers, interactive markers, and repetitions (van Zeeland & Schmitt, 2012). Furthermore, marginal words, such as *uh*, *ah*, *uhm*, *oh,* which are fillers in spoken communications, are usually more present in unscripted spoken discourse and does not

require learners to make form-meaning links to them. Comprehension when listening may, therefore, require a smaller vocabulary size than when reading.

Consequently, there seems to be a shortage of empirical research that supports the use of lexical coverage thresholds of 95% and 98% for listening comprehension. This also applies to watching videos, as studies assessing the vocabulary demands of movies and television programs have not investigated the necessary coverage rates for the particular discourse (Webb & Rodgers, 2009a, 2009b; Rodgers & Webb, 2011). Laufer and Ravenhorst-Kalovski (2010) suggest that the coverage necessary for comprehension depends on one's learning goals, and what suffices as "adequate" comprehension for the individual learner himself. They also suggest that with a dense lexical coverage of a discourse, the percentage of unknown words is so low that the words, most likely, will either be non-essential for understanding of the discourse or the meaning of the unknown words may be understood from the context. Based on the arguments presented, and in order to compare the results to previous research, coverage thresholds of 95% and 98% will be used for the purpose of this study.

Figure 2.1 summarizes the empirical research conducted on the vocabulary size needed to understand different discourses. It also shows the differences in required vocabulary size for listening to spoken discourse and reading written discourse.

Word families	Discourse/text type	Evidence
The most frequent 3,000	Conversation	Nation (2006), van Zeeland & Schmitt (2013)
	television	Webb & Rodgers (2009b), Rodgers and Webb (2011)
	films	Webb and Rodgers (2009b)
The most frequent 4,000	Academic spoken discourse	Dang and Webb (2014)
The most frequent 8-9,000	Written texts, e.g. newspapers and novels	Nation (2006)

Figure 2.1 Number of word families required for comprehension. Reproduced from *How Vocabulary is Learned* (p. 45) by S., Webb and P., Nation, 2017, Oxford: Oxford University Press, Copyright 2017 by Oxford University Press. Reproduced with permission.

2.5 Listening in a Multimedia Environment

This section intends to outline the concepts of multimedia and multimedia learning. It will first describe the theories making up the theoretical framework for the present study and related principles, and lastly, it will present recent research on L2 listening comprehension in a multimedia context.

2.5.1 Related theories

This section aims to outline the theories making up the theoretical framework for the present study. The theories that will be presented are the cognitive theory of multimedia learning (Mayer, 2001) and the cognitive load theory (Sweller, 1988). The present study will base its discussion on the theoretical foundation framed by these two theories. Hence a detailed description of the theories will be presented in the following section.

2.5.1.1 Cognitive theory of multimedia learning

According to Mayer (2014b), the cognitive theory of multimedia learning is extended from general cognitive theories of learning. The rationale for the theory of multimedia learning is

based on the claim that "people learn more deeply from words and pictures than from words alone" (Mayer, 2014a, p. 43), which later has become known as the multimedia principle (see 2.5.2.1).

The cognitive theory of multimedia learning suggests that people have three memory stores for processing information: the sensory memory, the working memory, and the long-term memory. The sensory memory holds "sensory copies" of incoming words and pictures but can only hold them for a very brief time. The working memory is used for holding and manipulating knowledge temporarily in "active consciousness" (Mayer, 2014b, p. 53) whereas long-term memory can hold large amounts of knowledge over long periods but can only activate this knowledge by bringing it into working memory (Mayer, 2014b)

Furthermore, the theory is based on three assumptions: the first assumption is "that the human information processing system includes dual channels for visual and verbal processing" (Mayer, 2014a, p. 43). Specifically, the theory suggest that the human mind has two channels of working memory, a distinction first introduced by Paivio (1986) and his dual-coding theory. The dual-coding theory seeks to explain how information is represented and processed in two cognitive systems, implying that when information is presented both verbally and visually, comprehension is more likely. This is clarified by describing how learners will make a correlation between two similar modes of information and, therefore, will have more opportunities to evoke the information (Jones, 2001). The dual-coding assumption suggests that learners construct verbal mental representations in the verbal system and visual mental representations in the visual system, where verbal information is in the form of spoken or written text and visual information is in the form of pictures, animations and so on. Other scholars, however, suggest a division between sensory modalities, where the information can be presented through the eyes (visual) or through the ears (verbal) (Baddeley, 1997). Written text is in this case processed in the visual channel. For the purpose of this study, the prior distinction will be used where written text is assumed to be processed in the verbal channel. Furthermore, it is important to note that the relation between the channels are not closed, and that learners are able to convert processed information in one channel to the other channel (Mayer, 2014b). An example of how the dual-coding theory may enhance comprehension is proposed by Kozma (1991), who suggests that, for learners with low prior knowledge, information that is dually-coded may help them fill in the blanks that are in other ways missing from their prior knowledge.

The second assumption of the cognitive theory of multimedia learning suggests that a reduced capacity for processing exists in each of these channels. This will force learners to

make decisions about what information they should focus on, as they cannot pay attention to everything due to limited capacity. The limited-capacity assumption will be further discussed within the cognitive load theory in the following section.

The third assumption is that "[people] actively engage in cognitive processing in order to construct a coherent mental representation of their experiences" (Mayer, 2014b, p. 50). For active learning to occur, Mayer (2014b) deems the following five cognitive processes to be essential: attending to relevant words for processing in verbal working memory, attending to relevant images for processing in visual working memory, organizing the selected words into a verbal model, organizing the selected images into a pictorial model and, last, integrating the word-based and picture-based representations with each other and with previous knowledge. These processes are illustrated in Figure 2.2.
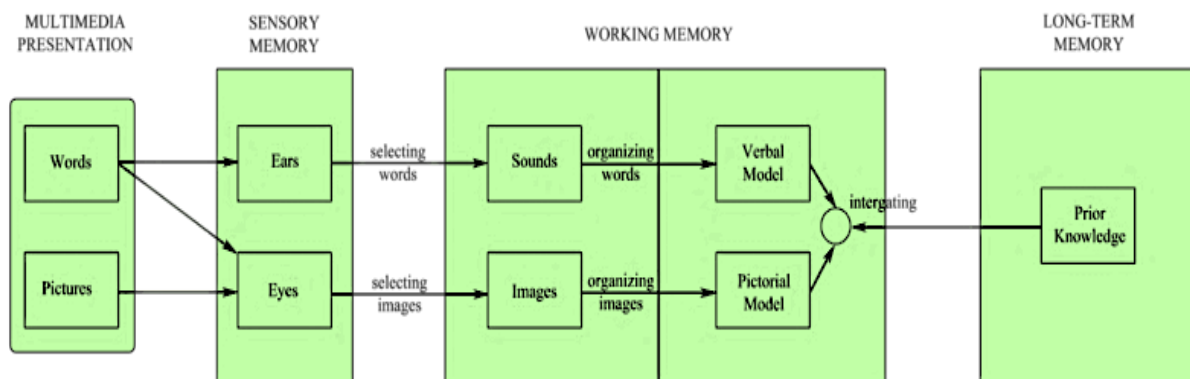
Figure 2.2 A model of the cognitive theory of multimedia learning. Reproduced from *The Cambridge Handbook of Multimedia Learning* (p. 52), R.E., Mayer (2014), New York: Cambridge University Press, Copyright 2005, 2014 by Cambridge University Press.

Thus, the cognitive theory of multimedia learning suggests that multimedia learning is a demanding process where a learner needs to select, organize, and integrate relevant words and images. This theory provides a theoretical foundation in the current study by suggesting a model of how multimedia input is processed. This will be used to discuss how input modes may have an impact on L2 learners' listening comprehension.

2.5.1.2 The cognitive load theory

The second theory within the present study's theoretical framework is the cognitive load theory. This theory also posits that working memory is limited and that long-term memory is unlimited in storing what is learned (Sweller et al., 2011). However, the theory emphasizes

how the working memory only has limitations when it is processing novel information "fed through the sensory [memory]" (Paas & Sweller, 2014, p. 34). Information that has been organized in long-term memory can be retrieved into working memory, and this information does not have the same limitations as information that is not yet organized in long-term memory. Consequently, long-term memory information vastly expands working memory information (Paas & Sweller, 2014)

When a learner encounters new information that must be learned, there may be too many elements to process simultaneously in working memory because working memory has limited capacity, which may cause a heavy cognitive load. Cognitive load refers, in this context, to the burden placed on working memory by the presented information (Paas & Sweller, 2014). The elements of information must be organized and combined with knowledge held in long-term memory, and if the elements that learners need to process are essential information, understanding cannot happen until it becomes possible to process all of the elements. Thus, "[w]hen knowledge acquisition has progressed to the point where all of the elements essential to understanding a topic can be processed in working memory, understanding has occurred" (Paas & Sweller, 2014, p. 36). Consequently, understanding is perceived as the skill where learners simultaneously process essential elements in working memory, which emphasizes the close relationship and interaction between working memory and long-term memory for understanding.

Paas and Sweller (2014) puts forward three categories of cognitive load: intrinsic, extraneous, and germane cognitive load. Intrinsic cognitive load is described as the cognitive load caused by interacting elements that are essential for solving the task and, thus, must be processed simultaneously. That is, a learner cannot learn one element without learning the interacting element at the same time. Extraneous cognitive load may happen when non-essential elements interact with essential elements, and require the learner to use working memory resources to process all of the elements – without acquiring knowledge. This type of cognitive load can be excluded by altering the instructional design of the multimedia text. Intrinsic cognitive load and extraneous cognitive load are additive and may together exceed working memory capacity if they are heavy loaded. Furthermore, the last category, the germane cognitive load, refers to a positive cognitive load where a learner uses more working memory resources to deal with intrinsic rather than extraneous cognitive load. For this to happen, the working memory resources needed to deal with extraneous cognitive load must be reduced to free working memory capacity (Paas & Sweller, 2014).

The cognitive load theory is, therefore, associated with learning of complicated cognitive tasks, where learners may be overwhelmed by the amount of elements that must be processed simultaneously before understanding can occur, which often imposes a heavy cognitive load on their working memory (Paas, van Gog, & Sweller, 2010). For the purpose of this study, watching videos found on YouTube may be regarded as such a complicated cognitive task, as videos often contain multiple interacting elements which the learner must process before comprehension can occur.

Consequently, the cognitive load theory provides an idea of how different uses of input modes in videos can impact the cognitive load in L2 learners and, thus, impact their listening comprehension. This theory is important in relation to the aim of the present study, as it may raise awareness of how designs of videos may be of help or hindrance to L2 learners' listening comprehension.

2.5.2 Related principles

On the premises of the previously discussed theory, I have chosen to introduce three principles which will be guiding the discussion of the results in the present study. The principles are limited to include the multimedia principle, the split-attention principle, and the redundancy principle.

2.5.2.1 The multimedia principle

The multimedia principle, proposed by Mayer (2001), suggests that learners learn more from the combination of visual and verbal input than when provided with text alone. The principle is consistent with the predictions from the previously mentioned theories of multimedia learning (Butcher, 2014). The multimedia principle also suggests that multimedia can enhance the linguistic input and increase the likelihood of learners paying attention to it (Plass and Jones 2005).

Several studies have investigated the effects of the multimedia principle. Nugent (1982) compared several combination of input modes and found that learners had a better recalling for stories when visual and auditory input were combined. He also found that visual, audio, and textual input containing the same information endorsed learning from each mode but was further increased when combining pictures with *either* text or audio. In addition, Lee, Lee, Liao, and Wang (2015) compared the effectiveness of foreign language learners' interpretive skills by giving one group of learners audio-visual aids in the form of video-clips,

while the other group did not receive these aids. They found that audio-visual aids were helpful in enhancing learners' listening comprehension, and that, in the listening test, audio-visual aids facilitated listening comprehension better than audio used alone. Brett (1997) investigated the comparative success rates on comprehension tasks and language recall through use of audio, video and multimedia, and found that higher levels of comprehension and language recall were achieved while listening in the multimedia environment. Guichon and McLornan (2008) conducted a pilot study investigating the effects of multimodality on comprehension using an audiovisual recording. They found increased comprehension when learners listened to a text in several modalities. Furthermore, research has found that the effectiveness of the multimedia principle varies with individual differences in spatial abilities, prior knowledge, and general learning abilities (Fletcher & Tobias, 2005). However, this is not within the scope of this thesis and will not be further dealt with.

In the present study, I will investigate how different input modes are combined in order to gain an understanding of how the multimedia principle may be applied in the selected videos in the corpus to enhance L2 listening comprehension.

2.5.2.2 The split-attention principle

The second principle which will be introduced is the split-attention principle. This principle suggests that presenting distinct sources of information in the same modality (e.g. verbal - such as audio and text) causes the learner to divide his attention, imposing a heavy load on cognitive processing which results in fewer resources for learning. Ayres and Sweller (2014) states that "in the design of instruction, including multimedia instruction, it is important to avoid materials that require learners to split their attention between, and mentally integrate, multiple sources of information" (p. 206). This extraneous cognitive load is likely to have a negative effect on learning, whereas eliminating the need to mentally integrate multiple sources of information reduces extraneous cognitive load and frees working memory resources available for processing. A split-attention effect may be caused when several essential elements of information are presented in disparate and not integrated ways, or when more than two sources are presented simultaneously. Split-attention may be limited by directing learners' attention to the relevant parts of information.

Schnotz (2014) problematizes the use of combining spoken text, written text, and pictures. He argues that "individuals do not learn better from pictures accompanied by spoken and written text" (p.92). The reason for this, he suggests, is that this combination will cause a split verbal attention because two elements will be processed in the same channel.

Furthermore, in a study conducted by Moreno and Mayer (1999), they found a split-attention effect even with subtle separations between verbal and visual elements. The group receiving separated format was forced to expend more cognitive resources to integrate the information for processing, which increased the extraneous cognitive load and led to a lower performance on the following test than the group receiving integrated information.

Furthermore, research indicates that individuals might read and listen in a different pace, and when L2 learners are forced to do both, it might interfere with the internal listening (reading) and external listening (what learners can hear), and eventually lead to a lower comprehension (Wetzel, Radtke, & Stern, 1993). In addition, Smith and Mosier (cited in Wetzel et. al, 1993) suggest that written text on the screen is more challenging to read than text in print, causing a 20 to 30 percent slower reading rate.

Mousavi et al. (1995) conducted six experiments investigating the split-attention principle and found that when learners are required to split their attention among multiple sources of information that must be mentally integrated before they can be understood, learning may be inhibited

Consequently, for the purpose of the present study which aims to examine the use of different input modes in videos, the split-attention principle suggests that if videos present their input modes in such a manner that learners will need to split their attention and mentally integrate the elements, comprehension may be hindered. This will be further discussed in Section 4.2.2.

2.5.2.3 The redundancy principle

Another application of presented theories is the redundancy principle. When the same information is presented concurrently in multiple modes or is unnecessarily detailed, redundancy may occur. Furthermore, the redundancy principle suggests that redundant material may interfere with rather than facilitate comprehension because a learner is forced to process redundant information, causing an increase in cognitive load (Kalyuga, Chandler, & Sweller, 1999). Adding redundant information with essential information increases the load imposed on working memory, which may interfere with learning due to heavy cognitive load, as suggested by cognitive load theory (see Section 2.5.1.2). Eliminating the redundant information removes the learner's need to organize multiple elements of information and, thus, reducing the cognitive load (Kalyuga, Chandler, & Sweller, 2004)

Scholars suggests that unnecessary or redundant on-screen text should be excluded or replaced by audio because viewers will typically try to read any readable text on the screen,

which will not support developments in listening comprehension abilities (Wetzel et al., 1993). For instance, Jamet and Le Bohec (2007) examined the redundancy effects obtained when spoken information was duplicated in writing during the learning of a multimedia document. They found that the duplication of information in the written mode led to cognitive overload and harmed learners' comprehension, which supports the redundancy principle proposed by the cognitive theories presented in this study.

Conclusively, the redundancy principle implies that, if videos present the same input in the same modality (verbal/visual), redundancy may occur and be of hindrance to learners' comprehension.

2.6 Chapter Summary

This chapter has presented the theoretical framework included in the present study and previous research related to the theories. The theoretical framework was based on the cognitive theory of multimedia learning and the cognitive load theory. This was chosen because the current study deals with teaching materials in the form of videos in a multimedia context and seeks to understand how this multimedia input is processed by L2 learners. The related principles discussed in this chapter have been limited to include the multimedia principle, the split-attention principle, and the redundancy principle. For the purpose of this study, these principles describe how different input modes may impact L2 learners' processing of multimedia input, and thus, may have an impact on their listening comprehension when they are exposed to videos found on YouTube in the classroom.

Furthermore, because viewing videos requires primarily listening skills, this chapter has outlined three processes involved when listening: decoding, comprehension, and interpretation. This has been done to better understand what processes are involved when L2 learners are listening to videos (see Section 2.3). As discussed in Section 2.3.1, an important feature involved when decoding input when listening is word knowledge. Hence, word knowledge is an essential part of the listening process and is an important aspect of the present study. If learners do not know the words, they will not be able to understand videos. Therefore, this study will investigate the vocabulary size needed for L2 learners to comprehend videos found on YouTube. The aim is that this will provide a better understanding of the lexical demands teachers place on their learners when presenting videos in the classroom.

L2 learners will probably never reach a native-like vocabulary size. Therefore, some words are more important to spend time learning learn than others. This is especially

important in the EFL classroom, as teachers have limited time to cover all parts of the subject curriculum. Furthermore, as discussed in Section 2.4.2, one way to decide what words are more expedient to learn than others is to look at how frequently certain words appear in a text. This is best done using Corpus Linguistics, which was discussed in Section 2.4.1, and which will be used as a method for assembling the present thesis' corpus.

Therefore, the current thesis seeks to investigate L2 learners' listening comprehension in a multimedia context in order to better understand the relationship between lexical knowledge, the use of input modes, and L2 learners' comprehension of videos used in the classroom. The hope is to raise teacher awareness to factors that may have an impact on the learners' listening comprehension when teachers use videos available on YouTube as classroom materials. In addition, this thesis hopes to raise awareness to what teachers need to consider when choosing such videos to use in their classrooms.

Conclusively, the present study investigates the following research questions:

1. What are the lexical demands placed on L2 learners when watching BBC News videos available on YouTube in the classroom?
2. How are different input modes presented in these videos?
   2a. How can the use of input modes in these videos impact listening comprehension?

3. Methods and Materials

The present study is a corpus-based study that has primarily used a quantitative method research design to gather and analyze the data. However, the method research design shows qualitative characteristics which will be described in the following sections. Three types of analyses were conducted in order to collect the data needed to answer the research questions presented above. This chapter provides a thorough description of the research framework for the current study and is separated into five sections. The first section will provide a description of materials. The second section will explain the choice of research methods and describe the research design. The third section will map out the data analysis process. The fourth section contains a discussion of the ethical issues, while the final section will discuss validity and reliability relevant for the present thesis.

3.1 Materials

For the present study, two sets of transcripts were developed from twenty BBC News videos concerning topics in the US, which were selected through purposeful sampling based on a set of criteria (see Section 3.1.3.2). The first set of transcripts comprised a small corpus of 9,692 tokens which were analyzed using computer software and frequency word family lists from the BNC/COCA word lists (Nation, 2017). The first set of transcripts contained words that were essential information for understanding the message of the videos. The second set of transcripts contained, in addition to essential information, on-screen text appearing in the background (non-essential information) and on-screen text with aural narration. The following section will provide a more detailed description of the materials used to conduct the present study.

3.1.1 The software

The instrument used in the present study to gather the quantitative data for the research was the AntWordProfiler software, created by Laurence Anthony (2014). The AntWordProfiler is a computer software which lists words according to frequency (the number of occurrences in a given corpus) and range (the distribution of words or other lexical items across a given corpus). The software is available for download from Laurence Anthony's website (www.laurenceanthony.net).

The software was chosen because it is a freeware available for everyone, and it has the ability to analyze several texts in batches which reduces the work load. Furthermore, the

software allows for alternative word lists to be uploaded in addition to its default lists. The software is recommended by other researchers, for example Nation (2014), for this type of use. According to Creswell (2014), the chosen instrument should have an acceptable scale of measurement and provide reliable and valid scores, which were factors I looked for when choosing the software for the present study. To check if the software provides reliable and valid scores, I ran the transcripts through a similar software called Lextutor[4], developed by Cobb This software does also list words in a text according to frequency. The results from the Lextutor were comparable to the results provided by the AntWordProfiler, which enhanced the notion that the software chosen for the current thesis was a reliable software to use for gathering my data.

3.1.2 The word lists

The word lists used for the analysis in this study were Nation's BNC/COCA word family lists. They are available for download from the Victoria University of Wellington's website (Nation, 2017). The lists consist of 29 word family lists based on the BNC and COCA corpus. Twenty five of the lists contain word families based on frequency. That is, the headwords of the first 1,000 word-level (*a*, *able*, *about*) are more frequent than those at the 3,000 word-level (*abandon*, *abort*, *abroad*), and so on. Each list has a fixed name, *basewrdx.txt*, where *x* is a number and refers to the name of the list. The four additional lists (basewrd31-34) are one, a list of proper names, two, a list of marginal words (including swear words, exclamations, and letters of the alphabet), three, a list of transparent compounds, and four, a list of abbreviations. Words less frequent than the 25,000 level were placed in an "off-list". Spelling errors, abbreviations, slang, contractions, and so on that did not belong to any list were placed in this list by the software. However, in order to limit the number of words placed in this list, the words were manually corrected, written out, or added to their appropriate list. Thus, a minimal number of words is placed in the off-list, and no level of frequency exists for this list (see Table 3.1).

All word stems are free forms, not bound forms, which means that they can stand alone (bound forms do not carry meaning, and, thus, do not constitute as separate words). The word lists contain compound words but do not include phrases. Phrases and multiword units will not be within the scope of this study (see Section 5.2). The lists use the word family as the unit of counting (see Section 2.4.2.1). The analysis of the videos using word families was

---

[4] Available from http://www.lextutor.ca/.

likely to provide an acceptable measurement of the vocabulary in the BBC News videos because watching these videos requires receptive knowledge of the words that are encountered rather than productive knowledge. If the word lists were created for productive purposes, as when speaking and writing, the lemma would be the most reasonable unit to use because each lemma takes different collocates and different grammatical patterns (Nation, 2006).

Furthermore, based on Bauer and Nation (1993), the criteria for the word families were set to Level 6, which includes all the affixes from levels 2 to 6 (with more than 80 derivational affixes). Researchers argue whether Level 6 is the appropriate level for L2 learners, as the number of affixes might be too high for their knowledge. This means that, even though the learner might know a word family, they might not know all the derivational affixes up to level 6. However, in order to be able to compare my findings to previous research, level 6 is used as the appropriate level in this study (see Section 5.2).

3.1.3 The videos

3.1.3.1 Survey

I initially wanted to conduct a study examining the lexical demands of YouTube videos that teachers actually used in the EFL classroom in Norway. In order to obtain this information, I decided to construct a survey. Using a mixed method research design, the survey had a combination of closed and open-ended questions for collecting data. This type of research design was chosen because I wanted to get a detailed understanding of the quantitative data collected from the closed questions, which was best done by also collecting qualitative data from open-ended questions. However, open-ended questions were in one way unavoidable in this survey because I wanted teachers to provide examples of YouTube videos that they actually used. Gathering such data would have been difficult with closed questions only, as it required teachers to write out their own examples in their own words. Additionally, closed-questions were used to gather information about the teachers' use of specific YouTube channels, in case they did not provide examples of specific videos (see Appendix I for an overview of the survey questions). Thus, I needed to integrate both quantitative and qualitative methods of data collection, which can be identified as mixed methods research (Clark & Ivankova, 2015).

According to Punch (2003), a survey needs to be piloted "to test for comprehension, clarity, ambiguity, and difficulty in responding to"(p.39). Thus, I piloted the survey twice –

first with a group of peer students, then with two teachers, to ensure that the survey worked on all computers, to see how long the survey would take for the respondents to finish, and that all the questions were comprehensible. I then reevaluated the survey based on their feedback.

I distributed the survey via e-mail to two lower secondary schools in each county in Norway, 36 schools in total. The information about the schools was found on www.skoleadresser.no and is public information. The survey remained active for a month and a reminder e-mail was sent to the schools after three weeks.

Unfortunately, only seven teachers replied. Of the seven respondents, four provided examples of YouTube videos they used in their classrooms, and these videos were so diverse that there would be no point in analyzing them for the sake of this study. Therefore, the survey did not provide me with enough information to conduct a study solely based on the YouTube videos teachers reported using in the classroom, which was the intention of the survey. However, the survey did give me valuable information of which YouTube channels teachers like to use in the classroom as most of the respondents reported using BBC's YouTube channel in their classroom (see Figure 3.1). Based on these results, I chose to analyze videos produced by BBC available on YouTube. Consequently, in the present study, the survey did not provide any specific answers to what I set out to investigate but gave me an idea for which videos I could use for the purpose of this study.
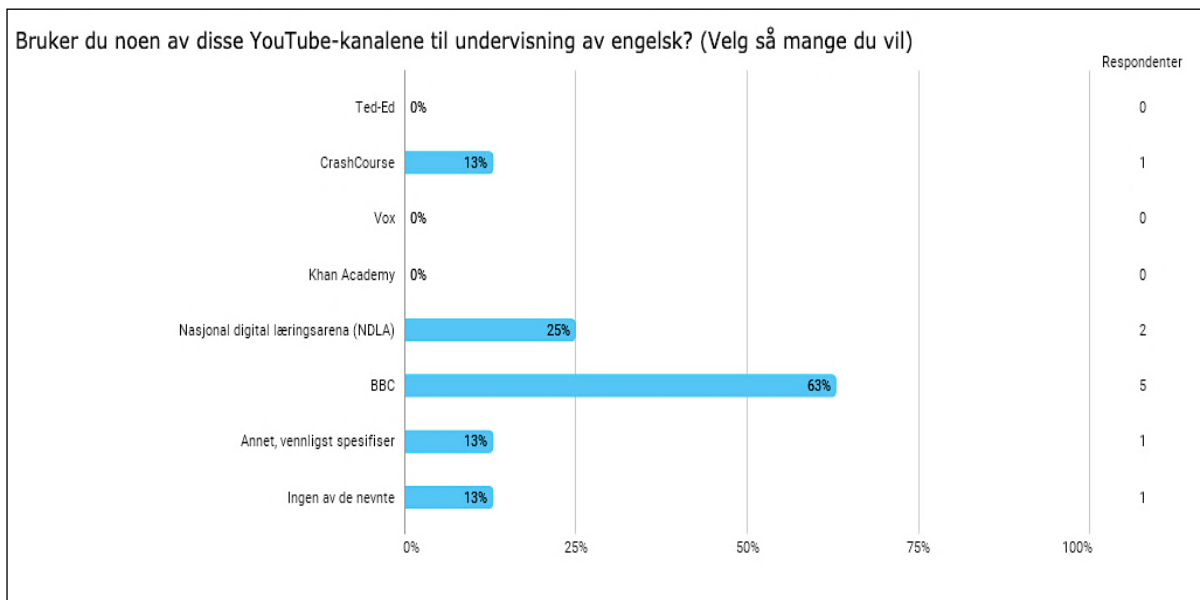

Figure 3.1 results from survey question about the use of YouTube channels

3.1.3.2 Criteria for choosing videos

Online content is available for anyone on the BBC website as well as on BBC's YouTube channels. BBC creates a variety of different videos related to various topics. In addition to its main channel on YouTube, BBC has different subchannels, such as BBC News, BBC Docs, and BBC Earth – of which BBC News has most videos available. Even though it is a British broadcaster, the BBC News videos are news reports about recent topics from all over the world.

The following criteria were set before selecting videos for the analysis: all videos had to be selected from the BBC News channel on YouTube, they had to be less than ten minutes long, they had to be suitable for children under the age of 16 – which meant that videos including swear words, bad language, and/or explicit content were not included. Being a teacher in training, all videos had to be videos I would be likely to use in the classroom myself. Lastly, the videos had to be related to American culture and society, that is, topics concerning politics, youths, gun-control, poverty, immigration, sports, or discrimination were preferable. The reason for selecting videos that concerned the US was that "being able to discuss and elaborate on the way people live and how they socialise in Great Britain, USA and other English-speaking countries and Norway" (The Norwegian Directorate for Education and Training, 2013, p. 9) are placed in the English Subject Curriculum as a competence aim after Year 10[5]. The BBC News channel on YouTube has several videos concerning American culture and society, which English teachers may use in their classrooms in relation to this competence aim.

Twenty videos were selected in total. They were between one and ten minutes long and had a total running time of 57 minutes (for further descriptions see Appendix II). All videos were available on the BBC News channel on YouTube. The videos I selected were centered around their content, where I only selected videos about the US. The videos contained topics such as politics, sports, gun-control, school-shootings, and immigration, topics that are, or were, a big part of the US news picture within the last two years, and, which might be relevant when discussing American culture and society with regards to the competence aim previously mentioned.

---

[5] It must be mentioned that the core curriculum is undergoing a restructuring, and that new subject curricula will be implemented from 2020. This also concerns the English subject curriculum. However, after looking at the proposed draft for the new curriculum, the competence aim: "being able to present and discuss current news from English-speaking countries" seems to still make these videos relevant for the EFL classrooms.

3.1.4 Transcripts

Two sets of transcripts were constructed from the videos selected. The first set included all spoken and written words necessary for understanding the message of the video. These transcripts comprised the corpus which was analyzed to determine the lexical coverage using the AntWordProfiler. The second set of transcripts, based on the same videos, was a more detailed extension of the first set, in that it also included words appearing as on-screen text in the background, and words presented as on-screen text with aural narration. That is, words a viewer would be exposed to, but were not essential for understanding the message. The second set of transcripts comprised the material for the second set of analyses.

3.1.4.1 Principles for transcripts

While Webb and Rodgers (2009a, 2009b) used transcripts downloaded from the Internet when assessing the lexical demands of movies and television programs, this was not available for the BBC News videos. Thus, in order to ensure that all videos were transcribed in the same manner, I had to develop my own set of principles prior to the transcription process. Thus, the following principles were followed when creating the transcripts: numbers were written out, that is, *2018* became *two thousand and eighteen*. Contractions, connected speech, and hyphenated words were conformed into their spellings in the BNC/COCA lists. For example, words such as *wanna, shoulda*, and *dunno* were changed to *want to*, *should have*, and *do not know* respectively. If the spellings had not been changed, the items would have been classified as being less frequent than the most frequent 25,000 word families and would have been placed in the off-list.

With regards to compound words, transparent compounds were split into two parts where each part was counted as a separate word, a solution used by Brysbaert, Stevens, Mandera, and Keuleers (2016). Compounds that were not transparent were added to the compound word list (baseword33) since splitting these words would create words in the off-list or change the core meaning of the word. These words accounted for 0.22% of the total tokens. Abbreviations classified as not in the lists were reclassified and put in the abbreviation-list (basewrds34), seven tokens in total (0.07%). Fillers and hesitations such as *oh*, *uhm*, *ah*, and so on were classified as marginal words and put in the marginal words-list (basewrd32), 31 tokens in total (0.32%). Stuttering and "half-repetitions" of words were written once, for example, "instead of critiq… critiquing us, come help us,"[6] the word

---

[6] Video 4, see Appendix II.

critiquing was only written once. This half-pronounced word would have been classified as off-list, because it is not a word.

Names and titles of speakers, interview objects, and place names that appeared on-screen as written text were transcribed. Names that were mentioned orally were also transcribed. Proper names found in the off-list were reclassified as proper nouns and put in the proper nouns-list (baseword31) to avoid tokens in the off-list.

However, there was one particular case for one of the videos where I was unable to hear a word the speaker said. I decided to count the word as a token but chose to manually place it in the off-list.

The AntWordProfiler software does not, unfortunately, distinguish between homographs or capital letters, which caused several problems when transcribing the videos. To deal with this, the following steps were taken:

- Acronyms that could be classified as anything else than an acronym were written out in full to avoid being counted as a more or less frequent word. For example, the acronym *US* was changed into *the United States* to avoid it being counted as *us*.

- A problem appeared with the proper noun *Trump*, as the American president per 2019, Donald Trump. The AntWordProfiler classified it as a verb (*to trump*) which belongs to the 6,000 word family level and not in the proper nouns list (basewrd31). This resulted in 21 tokens belonging to the 6,000 word family list, which should have been classified as proper nouns. In an attempt to limit this problem, I chose to manually reclassify *Trump* as a proper noun and place it in the proper nouns list as this provided a more accurate result. If these tokens had not been changed, the frequency levels needed for reaching 95% and 98% coverage would have been higher.

- There was one case where the collocation *to be frank* caused a problem because *frank* (as the adjective) was counted as *Frank* (the proper noun). *Frank* (adjective) was not found in any other list than the proper nouns-list, but *frankly* was found in the 2,000 level. If one assumes that *frank* belongs to the word family *frankly*, then *frank* would be reclassified as belonging to the most frequent 2,000 word families. I tried this adjustment to see if it changed the results; however, it did nothing to the coverage percentages in the analysis, so I left it as it was.

The accuracy of this study would have been higher if the software was able to distinguish homographs. However, the number of tokens is so small the overall analysis should not be affected by this.

3.1.5 The corpus

The corpus used for the first analysis was a small targeted corpus consisting of twenty BBC News transcripts. The videos were selected to form a corpus that was intended to represent BBC News videos available on YouTube. The corpus consisted of 9,692 tokens in total. BBC does not provide transcripts for all of its videos, which is why I transcribed them by hand following a set of principles (see Section 3.1.4.1). It is important to note that these transcripts might not accurately reflect the words in the videos, however, the transcripts should provide a consistent assessment of the vocabulary used in BBC News videos.

To date, relatively little is known about the lexical demands of YouTube videos used in the EFL classroom, which means that there are, to the best of my knowledge, no clear guidelines or previous studies to how one should analyze YouTube videos in this type of study (see Section 2.4.4.1). As mentioned, spoken words as well as written words that were essential to be able to understand the message of the video were included in the first set of transcripts and comprised the corpus used to gather data for the first analysis.

3.2 Research Methods and Design

This section aims to give a detailed description of the research methods and design chosen for the present study. Morgan (2017) states that "[e]very successful research project requires two things: a meaningful research question and an appropriate way to answer that question" (p. 20). Thus, for any empirical study, the purposes of the study will guide the choice of research method.

In the current thesis, the methods chosen were guided by the research questions. Both research questions follow the description of quantitative research questions as proposed by Cresswell (2014) who states that quantitative research questions need to be "specific, narrow, measurable, and observable" (p. 27). Thus, the research questions in the present study spawn the need for the quantitative methods. However, the research methods applied to the present study have both quantitative and qualitative features, because, as Dörnyei (2007) views it, the qualitative-quantitative distinction forms a continuum, not a dichotomy.

Three different analysis were conducted in the present study in order to answer the research questions. Following the description of the main characteristics of quantitative research presented by Dörnyei (2007), the present study has quantitative features in that it is centered around gathering numbers, the categorization is specified before the study, and that the data is analyzed using statistical procedure. In the first analysis, Corpus Linguistics (see section 2.4.1) has been used as a method to bring about a general analysis of the lexical

demand of the BBC News videos, and to provide the foundation for an investigation of questions concerning the broader discussion of listening comprehension. The software in this present study was used to generate objective, numeric data of general patterns of frequency levels across the corpus in order to determine the coverage percentages. For the second analysis, a quantitative method was applied by manually collecting numeric data from pre-determined categories, and later, the data was analyzed using statistical procedures. On the basis of the data collected from the second analysis, a third analysis was conducted. The third analysis had quantitative characteristics in that the collected data was numeric data.

The current thesis has qualitative characteristics due to the corpus being assembled through purposeful sampling (Cresswell, 2014), and that it contained a relatively small number of tokens. As previously stated, the corpus examined in this present study consisted of transcripts from 20 videos selected from the BBC News' YouTube channel with a total of 9,692 tokens. Purposeful sampling is a technique commonly used in qualitative research for the deliberate collection of specific examples in order to gain deeper information about the research topic (Patton, 2002). Thus, purposeful sampling seemed to be the most appropriate technique for gathering data in the current study. Morgan (2017) states that

> qualitative studies thus tend to rely on a careful process of purposive selection to locate the data sources that are most relevant to the research topic. In these cases, the preliminary use of even a small quantitative study can provide important resources for targeting the most productive or theoretically relevant sources for [the] qualitative data (p. 17-18).

The samples used to form this thesis' corpus were intentionally selected to help understand the use of vocabulary and input modes in these videos. A purposeful sampling may provide useful information about the samples studied (Cresswell, 2014). Furthermore, qualitative characteristics are found in the third analysis, which was conducted based on the results from the second analysis. A qualitative method was applied to this analysis in order to get a deeper understanding of the numeric data from the quantitative analysis.

3.3 Analyses

Three analyses were conducted in order to gather the data for this thesis. The following section will describe the procedures for the analyses.

3.3.1 Lexical coverage

In order to answer the first research question of how lexically demanding BBC News videos are for learners viewing them, a quantitative method using the software AntWordProfiler was applied to the small corpus previously described in Section 3.1.5.

Previous studies done by Webb and Rodgers (2009a, 2009b), who assessed the lexical demand of TV-shows and movies were used as the inspiration for the choice of methods and materials for the present study. However, the present study used Nation's (2017) updated word family list based on the BNC and COCA corpora (see Section 3.1.2), while Webb and Rodgers (2009a, 2009b) used Nation's word lists from 2006, based on the BNC corpus only. Nation's (2017) word lists seemed to be more representative for the vocabulary used in the BBC News videos than the 2006 word list, as these videos contain both American and British English language variations, as well as more spoken text samples.

The twenty transcripts were run through the computer software, AntWordProfiler. The software generated the results in a table, where tokens were put in frequency levels starting from level one to level 32, as illustrated below in Table 3.1

Table 3.1 Statistical table generated by the AntWordProfiler

Statistics										
Level	File	Token	Token%	Cumtoken%	Type	Type%	Cumtype%	Group	Group%	Cumgroup%
1	basewrd1.txt									
2	basewrd2.txt									
3	basewrd3.txt									
4	basewrd4.txt									
5	basewrd5.txt									
6	basewrd6.txt									
7	basewrd7.txt									
8	basewrd8.txt									
9	basewrd9.txt									
10	basewrd10.txt									
11	basewrd11.txt									
12	basewrd12.txt									
13	basewrd13.txt									
16	basewrd16.txt									
25	basewrd25.txt									
26	basewrd31.txt									
27	basewrd32.txt									
28	basewrd33.txt									
29	basewrd34.txt									
0[?]	-									
TOTAL:										

[?]Level 0 has no list, see Section 3.1.2.

The software does not count proper nouns and marginal words as belonging to level 1, and, thus, the tokens belonging to these two lists had to be manually added to the first 1,000 level and the coverage percentages had to be recalculated. The AntWordProfiler has the ability to analyze up to 35 files at the same time, which made it possible to analyze all twenty transcripts in one batch. This way, the software generated the cumulative coverage of the total corpus. Furthermore, every transcript was also analyzed individually to make sure the analysis was accurate, and to generate the cumulative coverage of each separate video so that a

comparison between the total corpus and individual videos could be made. The data generated from the AntWordProfiler were collected in three tables showing the numbers of tokens and numbers of word families (see Table 4.1, Table 4.2, and Table 4.3).

3.3.2 Input modes

To answer the second set of research question of how different input modes were presented in the videos, a quantitative method was applied to collect the data.

A second set of transcripts was created from the twenty BBC News videos by transcribing every word appearing in any mode in the video[7]. In detail, this means that every word that a viewer would be able to read or hear during the video was transcribed, independently of the message of the video. Words appearing on-screen but did not stay long enough for me to read them were excluded, as other viewers most likely would not be able to read them as well.

Every word was placed into pre-determined categories of modes. The categories were chosen based on how words appeared in the videos, which I had observed during the process of creating the first transcripts. I followed primarily a division proposed by Mayer (2014c), who separates between on-screen text and spoken text for the verbal modality, and dynamic and static pictures for the visual modality. In addition, I included a division between essential and non-essential elements as proposed by Paas and Sweller (2014). The categories I chose were: spoken text (red), on-screen text (blue), non-essential on-screen text (green), or on-screen text with aural narration (purple). I used colors to categorize each word (see Figure 3.1 for an example of words marked with colors). The first three categories are words presented in single-mode (one element processed in one channel), while the last category contains words presented in dual-mode (two elements processed in one channel). For each transcript, all words were manually summarized and placed in a table according to category. Percentages of each category were calculated and placed next to the category (see Table 4.4).

---

[7] Note that words presented as written text with aural narration mode were transcribed twice and categorized as both red (spoken mode) and purple (dual mode). See Figure 3.2 for an example of this. Consequently, this affects these categories' total tokens and percentages for each video.

Figure 3.2 Example-text of how words were categorized as presentation modes using colors

3.3.2.1 Frequency analysis of on-screen text with narration

A third analysis was conducted on the basis of the results from the second analysis. Words belonging to the purple category (on-screen text with aural narration) were collected to form an individual transcript. The transcript was analyzed using the AntWordProfiler to observe what type of frequency-words were presented in dual-mode. The results are presented in Table 4.5 (see Section 4.1.3).

3.4 Ethical Issues

The current study does not contain any type of personal data and is not subject to notification requirements given by the Norwegian Social Science Data Services (NSD). However, when carrying out any type of research, ethical issues will always need to be taken into consideration. In this thesis, ethical issues are concerned with the use of videos for conducting research. The videos are considered works of others and need to be treated with respect. This entails that any type of analysis using these materials needs to be performed fairly. Furthermore, all data collected from the use of works of others needs to be "reported honestly, without changing or altering the findings to satisfy certain predictions" (Creswell, 2014, p. 38). The results should be presented in a sensible, unbiased way and any reference to other people's work must be cited accurately (Creswell, 2014).

Thus, I have done my best to make sure that my findings reflect an honest review of the data and information collected when conducting this research. I have also tried to provide a nuanced discussion of the findings to show respect for those that have worked to produce these video materials. I have, to the best of my ability, tried to stay as objective as possible when sampling videos for the corpus. Therefore, the purposeful sampling was done in a

systematic and objective manner in order to ensure that each video was not chosen based on my personal feelings. Choosing videos following a set of criteria, means excluding others. Thus, I have tried to not be affected by the content of the video, looking only at the title and the time of the video. I have not chosen videos based on how difficult or easy they were to transcribe or if the content of the video was exciting or not.

3.5 Reliability and Validity

In the present study, reliability and validity is closely linked to the choice of materials, especially the instruments that measures variables and how they provide reliable data. Reliability concerns the score that the instruments provide, and if these scores are stable, consistent, and reproducible. Validity relates to the instrument used, and if it actually measures what it intended to. Consequently, reliability and validity are connected in a complex manner, where they sometimes coincide, and sometimes disagree. However, according to Creswell (2014), "if a score is not reliable, it is not valid" (p. 177).

In the present study, the software AntWordProfiler was used to gather the numeric data. To test for its reliability, the corpus was analyzed several times with this program. Each transcript was also analyzed individually to see if the scores were stable. Additionally, as mentioned in Section 3.1.1, the transcripts were also cross-checked using Lextutor to test for reliability. I concluded that the AntWordProfiler proved to offer reliable scores.

For Corpus Linguistic research, the issue of representativeness of the corpus is important when accounting for reliability. In this current thesis, I have tried to provide a representative sample corpus of BBC News videos concerning topics from the US. However, as transcripts were not available for BBC News videos, the process of creating transcriptions proved to be very time consuming. Consequently, I was only able to produce twenty transcripts of BBC News videos for the corpus within the time restrictions I had for this study. Thus, the corpus can be considered too small to be generalized for every BBC News video available on YouTube. Although the corpus represents a small sample of BBC News videos with topics from the US, the study may provide an indication of the language use in these specific videos.

Lastly, being explicit about how the corpus is constructed is very important (Tognini-Bonelli, 2001). Consequently, I have tried to be as transparent about the process of assembling the corpus as possible, and I have reported my steps as honest and open as possible. I have provided a detailed description of the processes involved, and the choices I have made. Thus, the validity of the current research is dependent on reliable data which I

have, to the best of my ability, tried to ensure by being thorough, systematic, objective, and open about the data collection process throughout this project.

4. Findings and Discussion

The following chapter will present and discuss the findings for the present study. The chapter is divided into five sections. The first section presents the results from the first analysis, which seeks to answer the first research question of how lexically demanding BBC videos are on L2 learners. The second section illustrates the results from the second set of analyses in answer to the second set of research questions, which asked how different input modes are used in these videos, and how this can impact listening comprehension. The third section provides a discussion of the findings from the first analysis, while the fourth section will discuss the findings from the second set of analyses. Finally, a general discussion of the findings will be provided in the fifth section.

4.1 Findings

The following section will present the results from the analyses conducted in the current study. The findings from the first analysis, which examined the word frequency of the corpus in order to answer the first research question, will be presented first, then the findings from the second set of analyses, which investigated the use of different output modes in order to answer the second set of research questions, will be presented.

4.1.1 Lexical coverage

The first analysis examined at what frequency level 95% and 98% lexical coverage were reached for the selected BBC News videos in the corpus. The coverage thresholds were chosen based on previous studies (Adolphs & Schmitt, 2003; Laufer & Sim, 1985; Nation & Waring, 1997; van Zeeland & Schmitt, 2012), and similar studies (Webb & Rodgers, 2009a, 2009b; Rodgers & Webb, 2010).

Table 4.1 illustrates the raw figures and percentages of the total tokens and word families at each word level for the total tokens (see 3.3.1).

Table 4.1 Tokens, types, and word families at each word level for the total corpus

Word lists	Tokens		Word families
	Raw	%	
1,000	7,817	80.65	643
2,000	739	7.62	282
3,000	389	4.01	203
4,000	84	0.87	56
5,000	59	0.60	48
6,000	38	0.39	23
7,000	23	0.24	17
8,000	7	0.07	7
9,000	12	0.12	11
10,000	1	0.01	1
11,000	12	0.12	4
12,000	6	0.06	4
13,000	2	0.02	2
16,000	2	0.02	2
25,000	1	0.01	1
Proper nouns	439	4.53	175
Marginal words	31	0.32	4
Compounds	21	0.22	15
Abbreviations	7	0.07	5
Off-list	2	0.02	?[a]
Total	9,692		1,505

[a]The AntWordProfiler software is unable to calculate word families for words in the off-list.

Table 4.1 shows that the first 1,000 word families comprised most of the tokens and word families for the total corpus, with the number of each consistently decreasing as the word-frequency decreases, following Zipf's law (see Section 2.4.2). The most frequent 1,000 word families constituted 7,817 of the running words (80.65%). In contrast, the second set of 1,000 word families formed only 739 running words (7.62%). Only 0.39% of the running words are accounted for by the sixth set of 1,000 word families. Moreover, Table 4.1 also shows that 0.04% of the words were put in the off-list, which is a list containing words that are less frequent than the 25,000 word family level or non-words (see Section 3.1.2). The corpus used

9,692 words in total. Furthermore, illustrated in Table 4.1, the third highest number of tokens consisted of proper nouns (4.53%), while the number of marginal words made up 0.32% of the total tokens for the corpus.

As mentioned in Section 2.4.4.1, Nation (2006) suggested that proper nouns and marginal words should be added to the first 1,000 word families. Hence, Table 4.2 shows the cumulative coverage of the total corpus, with and without proper nouns and marginal words added to the most frequent 1,000 word families.

Table 4.2 Cumulative coverage, with and without proper nouns and marginal words for the total corpus

Word list	Coverage without proper nouns and marginal words	Coverage including proper nouns and marginal words
1,000	80.65	85.25
2,000	88.27	92.87
3,000	92.28	96.88[a]
4,000	93.15	97.75
5,000	93.76	98.36[b]
6,000	94.15	98.75
7,000	94.39	98.99
8,000	94.46	99.06
9,000	94.58	99.18
10,000	94.59	99.19
11,000	94.71	99.31
12,000	94.77	99.37
13,000	94.79	99.39
16,000	94.81	99.41
25,000	94.82	99.42
Compounds	95.04[a]	99.64
Abbreviations	95.11	99.71
Proper nouns	4.53	
Marginal words	0.07	

[a]Reaching 95% coverage

[b]Reaching 98% coverage

With knowledge of proper nouns and marginal words, Table 4.2 shows that a vocabulary size of 3,000 word families is sufficient for reaching 95% coverage of the total corpus. Furthermore, a vocabulary size of 5,000 word families, including knowledge of proper nouns and marginal words, will provide coverage of 98.36% of the total corpus.

Frequency levels where 95% and 98% coverage of each video were reached, including knowledge of proper nouns and marginal words are shown in Table 4.3. For an overview of the videos' titles see Appendix II.

Table 4.3 Vocabulary size for 95% and 98% coverage, individual video

Video	95%	98%
1	3,000	5,000
2	3,000	4,000
3	3,000	4,000
4	3,000	4,000
5	3,000	6,000
6	2,000	4,000
7	3,000	7,000
8	-	3,000
9	3,000	4,000
10	2,000	5,000
11	2,000	3,000
12	3,000	5,000
13	3,000	4,000
14	3,000	6,000
15	2,000	3,000
16	3,000	4,000
17	3,000	5,000
18	3,000	5,000
19	4,000	6,000
20	-	3,000

As Table 4.3 illustrates, for 13 out of 20 videos (65%), 95% coverage were reached with a vocabulary size of 3,000 word families for listening comprehension. Four videos (20%)

reached 95% coverage with knowledge of 2,000 words families, while for one video (5%) a vocabulary size of 4,000 word families were needed to reach 95% coverage.

Knowledge of 3,000 word families resulted in 98% coverage for two videos (10%), while a vocabulary size of 4,000 word families was needed for to reach 98% coverage for seven videos (35%). A vocabulary size of 5,000 word families was necessary to reach 98% coverage for five of the videos (25%), while for three videos (15%) a vocabulary size of 6,000 word families was necessary. For one video (5%), 98% coverage was reached with a vocabulary size of 7,000 word.

For the majority of the videos in this corpus, a vocabulary size of 3,000 word families, including knowledge of proper nouns and marginal words, was necessary to reach 95% coverage. Furthermore, a vocabulary size of 5,000 word families was needed to reach 98% coverage.

4.1.2 Input modes

While the results from the first analysis showed the lexical demands of the videos in the corpus, the second analysis investigated how information was presented in terms of input modes (see Section 1.4.2) for each video. This section puts forward the findings from the second analysis. Words categorized according to input mode (see 3.3.2) are illustrated in Table 4.4.

Table 4.4 Words categorized by type of input mode per video.

Video	Spoken	%	On-screen Text	%	On-screen text in the background	%	On-screen text w/ aural narration	%	Total tokens
1	50	16,7	238	79,3	12	4,0			300
2	1613	93,0	32	1,8	9	0,5	81	4,7	1735
3	307	70,3	58	13,3	21	4,8	51	11,7	437
4	381	98,2	7	1,8					388
5	283	79,5	15	4,2			58	16,3	356
6	501	92,4	38	7,0	3	0,6			542
7	323	80,5	61	15,2	17	4,2			401
8	150	76,5	46	23,5					196
9	375	76,8	98	20,1			15	3,1	488
10	495	80,0	120	19,4			4	0,6	619
11	306	95,9					13	4,1	319
12	511	96,1	21	3,9					532
13	670	84,3	81	10,2	4	0,5	40	5,0	795
14	378	66,8	90	15,9			98	17,3	566
15	133	60,7	86	39,3					219
16	530	74,9	152	21,5	9	1,3	17	2,4	708
17	231	90,9	12	4,7	11	4,3			254
18	290	74,7	60	15,5	5	1,3	33	8,5	388
19	395	98,0			8	2,0			403
20	410	97,2	10	2,4	2	0,5			422
Total time								Total tokens	10068

Table 4.4 shows that, while all videos used spoken mode, 19 out of 20 videos (95%) presented majority of their words in the form of spoken text mode. Only one video (5%) made use of more on-screen text mode than spoken text mode. Eighteen out of 20 videos (90%) presented words as both spoken and on-screen text modes but at different times in the videos.

Two videos (10%) did not present words from the on-screen text mode category. Additionally, 10 out of 20 (50%) videos presented words as on-screen text with aural narration mode, where the same words were presented in two modes at the same time (see Section 1.4.2). These videos presented 0.4% - 17.3% of their total words in this mode. Furthermore, 11 videos (55%) presented words in on-screen text mode that were non-essential, i.e. words appearing on signs, maps, or newspapers which are not necessary for comprehension of the video. Four of these videos (20%) had more than 4% of words presented this way.

As Table 4.4 illustrates, majority of the videos presented most of their words through spoken input mode, while on-screen text was the second most used input mode. Half of the videos made use of dual-mode by presenting words as on-screen text with aural narration. Only 11 videos had non-essential on-screen text appear in the background.

4.1.3 Frequency of words presented in dual-mode

Table 4.5 puts forward the results from the third analysis, which was conducted in order to examine the frequency of words in the on-screen text with aural narration category (dual-mode). The results show the numbers and percentages of tokens and word families presented in dual-mode for every frequency level, as well as numbers of proper nouns and compound words.

Table 4.5 Frequency levels of words presented in dual-mode, all videos.

Word lists	Tokens		Word families
	Raw	%	
1,000	331	77.34	136
2,000	45	10.56	23
3,000	11	2.58	8
4,000	2	0.47	2
5,000	3	0.70	3
6,000	8	1.88	3
7,000	5	1.17	5
Proper nouns	21	4.93	19
Compounds	2	0.47	2
Total	428		201

The results from Table 4.5 show that a total of 428 tokens were presented as on-screen text with aural narration. No words required knowledge of more than 7,000 word families. The highest number of words belonged to the first 1,000 word family level (77.2%). After the 3,000 level, the number of tokens decreased drastically, and no level after the 3,000 level had more than 10 tokens. Between the 4,000 level and 7,000 level there were only 18 tokens in total. Table 4.5 also shows that 21 tokens (4.9%) were proper nouns, while two words (0.47%) were compound words.

4.2 Discussion

This section will provide a detailed discussion of the findings from the analyses conducted in the present thesis. The results will be discussed in light of the previously mentioned theoretical framework (see Chapter 2). The section is divided into three sub-sections. First, a discussion of the findings from the first analysis related to lexical coverage will be provided. Second, the findings from the second set of analyses, which examined different uses of input modes, will be discussed. Finally, a general discussion of the findings will be presented.

4.2.1 Lexical coverage

This section takes a closer look at the findings related to the first research question, which seeks to answer what lexical demands teachers place on L2 learners when using videos found on YouTube as teaching materials. The first analysis examined the lexical coverage rates and expected vocabulary size. The results show that, with knowledge of proper nouns and marginal words, a vocabulary size of 3,000 word families is sufficient for L2 learners to reach 95% coverage of the total corpus. This means that if L2 learners have knowledge of at least 3,000 word families, one out of 20 words will be unknown. Furthermore, a vocabulary size of 5,000 word families, including proper nouns and marginal words, is necessary to reach 98% coverage of the total corpus. In this case, one out of 50 words will be unknown to the L2 learner.

Additionally, knowledge of 6,000-7,000 word families is suggested for reaching 98% coverage. These figures are supported by studies conducted by Webb and Rodgers (2009a, 2009b), who assessed the lexical coverage in movies and television programs. They found that the vocabulary size necessary for reaching 98% coverage of movies was 6,000 word families, while a vocabulary size of 7,000 word families was necessary for television programs. Their findings suggest that movies are slightly more demanding than television programs. In the current study, Table 4.2 show that knowledge of 5,000 word families is the vocabulary size necessary to reach 98% coverage of the total corpus. Thus, the results from the first analysis adhere to the figures found in previous research, but my results indicate that the videos in the corpus are slightly less demanding for reaching 98% coverage. It is important to note, however, that these results cannot be directly compared as the present study is not a replication of Webb and Rodgers' studies. Though their studies have served as the main inspiration for the methods used in the coverage analysis (such as the choice of word lists and software), there are fundamental differences in the size of the corpora used and type of data used for the analysis. This is because transcripts were unavailable for the selected videos in the present study, while Webb and Rodgers downloaded manuscripts online. The figures found in the present study supports the notion that spoken discourse is slightly less demanding than written text, as proposed by Nation (2006) who suggested knowledge of 6,000-7,000 families for spoken text and 8,000-9,00 for written text (see Section 2.4.4.1). It was also suggested that unscripted spoken text was less demanding than scripted spoken text, which may explain why the figures found in the current study are slightly less than Webb and Rodger's (2009a, 2009b) who examined exclusively scripted text.

The results presented in Table 4.2 also show that, without knowledge of proper nouns and marginal words, the vocabulary necessary to reach 95% is beyond the knowledge of the most frequent 25,000 word families. This is due to the fact that 4.53% of the total corpus consists of proper nouns. This finding illustrates the importance for L2 learners to be able to recognize words as proper nouns. As was discussed in Section 2.4.4.1, scholars (e.g. Hu & Nation, 2000) suggest that proper nouns should be considered known. Thus, I have followed this assumption and added proper nouns to the first 1,000 word families. However, for L2 learners, and particularly low-proficiency L2 learners, the situation might not be so straight forward. L2 learners need to be taught how to recognize names of people and places that are different from their L1 language, which is especially important when meeting homonyms in a text. For instance, *Hunter* (proper name) and *hunter* (noun), where the L2 learner will need to be able to distinguish between the two.

Additionally, the relatively high number of proper nouns is consistent with Nation's (2013) assumption that news texts will have a higher percentage of proper nouns than, for instance, novels. News videos usually present proper names as on-screen text simultaneously with the appearance of the interview object or reporter, which may be easily recognized by L2 learners (if one assumes the learners are familiar to news videos in their L1 language). In cases where this does not happen, they may have difficulties with recognizing proper names and this may hinder their comprehension, as proposed by the findings in the present study. As the results illustrate, if names of people and places are not assumed to be known, then 95% coverage will be beyond knowledge of the most frequent 25,000 word families. However, proficient L2 learners will probably be able to recognize names, as it is very unlikely that learners will have knowledge of 25,000 word families, but not be able to recognize names of people and places. Especially when considering they are very easy to learn (Nation, 2013). Nevertheless, it is important for L2 teachers to be aware that not knowing how to recognize places and names will cause comprehension issues for L2 learners when viewing videos in the classroom.

The results reveal that two tokens (0.4%) were placed in the off-list. This is important to be aware of, because the higher the number in the off-list, the more unreliable the results will be. One token were placed in this list purposefully, as it was a word I was unable to understand (see Section 3.1.4.1). Another token belonging to the off-list was the word *denuclearization* which is too infrequent to belong to the most frequent 25,000 word families. This is a word most L2 learners, even some native-speakers (see Section 2.4.4), are unable to understand without assistance (e.g. looking up words). However, some learners might know

one part of the word, such as nuclear, and use their knowledge of common and regular word-building devices in English to infer the meaning of the word (see Section 2.4.2.1). Nevertheless, as there are so few tokens in the off-list, this should not provide problems for the overall comprehension of the videos, and, thus, not affect the results of the study.

Conclusively, the results suggest that the minimum receptive vocabulary size necessary for understanding the total corpus should be 3,000 word families. Looking at studies conducted in Nordic countries (see Section 2.4.4), researchers found that only half of Danish and Icelandic L2 learners had a receptive vocabulary size of the most frequent 2,000 words. If one assumes that the situation in Norway is similar, reaching 95% coverage for the total corpus may not be realistic for all Norwegian L2 learners as they will need to have knowledge of around 1,000 word families more than the suggested vocabulary size of 2,000 word families. If 98% coverage is necessary for adequate listening comprehension, as proposed by several scholars (see Section 2.4.4.1), then the target vocabulary size may be too difficult a task for many L2 learners. However, more research is needed in this area, as there are still uncertainties of how large a vocabulary size Norwegian L2 learners have at the lower and upper secondary levels of education.

Furthermore, as Table 4.3 puts forward, there are individual differences between the videos in the corpus when it comes to lexical demands. For instance, while more than half of the videos (65%) required a vocabulary size of 3,000 word families for 95% coverage, four videos reached this coverage with knowledge of 2,000 words families, and one video reached it with a vocabulary size of 4,000 word families. These results reveal that there is a difference between understanding individual videos versus understanding the total corpus. Certain videos, then, might be more suitable for Norwegian L2 learners than others. In addition, this further enhances the notion that teachers need to be aware of how videos may impact L2 learners' listening comprehension. There is a huge difference between selecting a video that requires a learner to have knowledge of at least 3,000 word families and selecting a video where a vocabulary size of 2,000 word families will suffice. Thus, when teachers select a video based on topic and not the appropriate vocabulary level, it may have a huge impact on the number of learners who will be able to understand the video.

To sum up, while there are individual differences between the videos in the corpus, the findings from the first analysis show that a vocabulary size of 3,000 word families, including proper nouns and marginal words, are necessary to reach 95% coverage of the complete corpus. To reach 98% coverage of the total corpus, a vocabulary size of 5,000 word families will be sufficient.

4.2.2 Input modes

This section will provide a discussion of the findings from the second of analysis, which examined how words are presented in the videos. The findings will be discussed within the theoretical framework of cognitive multimedia learning theory and the cognitive load theory, and three related principles.

The second analysis was conducted in order to answer the second set of research questions, which asked how different input modes were used in the videos, and how this can impact L2 learners' listening comprehension. The results from the analysis show that the majority of the videos use spoken text as the primary input mode, while one video used primarily on-screen text modes. In addition, 18 out of 20 videos (90%) switched between the use of spoken text and the use of on-screen text. The findings show that when L2 learners watch these videos, they will not only be needing well-developed listening abilities, but they will also need to use their reading skills. Thus, presenting words as spoken input mode requires knowledge of the words' spoken form, while presenting words through on-screen text mode requires knowledge of the words' written form. L2 learners will, therefore, need to have knowledge of both spoken and written forms of words, and they will need to be able to recognize words when it is heard aurally and recognize words when they are seen on-screen.

The results, as illustrated in Table 4.4, show an inconsistent use of input modes in the videos. That is, there are no general patterns of how different modes are used between the videos in the corpus. The results also show that the majority of input modes in these videos are presented as single-modes. That is, one verbal mode is combined with one visual mode. As Mayer's (2001) cognitive theory of multimedia learning (see Section 2.5.2.1) suggests, learners possess dual-channels for processing information, a verbal channel for processing e.g. spoken words and written words, and a visual channel for processing e.g. pictures, diagrams, and animations. Furthermore, the multimedia principle suggests that learning is enhanced when verbal input is combined with visual input because it maximizes the resources available for both channels in the working memory (see Section 2.5.2.1). All the videos in the corpus provide visual input, and as the results presented in Table 4.4 illustrate, all the videos combine visual input with verbal input presented as either spoken text or written on-screen text modes. The multimedia principle suggests that L2 learners may find support in the visual input that the videos provide to better comprehend words they do not understand. Thus, when combined with either spoken text or written text, comprehension may be increased.

In relation to the current study, the results from the second analysis show that the majority of words found in the videos are spoken words. These words will be processed in the learner's verbal channel simultaneously as the visual input from the videos will be processed in the visual channel. In addition, 18 videos (90%) presented words as on-screen text in combination with visual input, which are words that will also be processed by the learners in their verbal channel. Thus, in relation to research question 2a, the combination of verbal and visual input in these videos may enhance L2 learners' listening comprehension, as proposed by the multimedia principle. How exactly the visual input may provide contextual support to the verbal input is not within the scope of this thesis (see Section 5.5 for suggestions for further research).

In addition, the cognitive load theory (see Section 2.5.1.2) suggests that when elements must be processed in the same channel, it can impose a heavy cognitive load, which in the worst case, can impair L2 learners' listening comprehension. In the present thesis, Table 4.4 reveals that half of the videos present words in dual-mode where on-screen text is presented simultaneously with aural narration. This means that the same word is being processed twice in the same channel, because the learner is exposed to the word in two different modes – spoken text and on-screen text. Both will be processed in the verbal channel, which may cause a heavy cognitive load due to limited capacity in the channel.

For the results found in the second analysis, the extraneous load category, proposed by the cognitive load theory (Sweller, 1988) seems most relevant. Extraneous cognitive load may be imposed when non-essential elements interact with essential elements (see Section 2.5.1.2). As Table 4.4 illustrates, 11 videos (55%) included words appearing on signs, maps, or newspapers, which are words not necessary for comprehension of the video. Thus, these words can be regarded as non-essential elements, as they could be excluded from the video, and the video would still make sense. Four of these videos (20%) had more than 4% of words presented this way, which are words that learners must process simultaneously. When L2 learners watch these videos that contain non-essential words, the non-essential elements will require the learner to use working memory resources to process them. Consequently, L2 learners may experience an extraneous cognitive load when they must process non-essential and essential words simultaneously, because the learner is only able to hold a limited number of elements in working memory. This type of cognitive load can be excluded by altering the instructional design (Sweller et al., 2011).

In addition, as the split-attention principle suggests, if these input modes are not coordinated, a split-attention effect may occur where the learner will need to use his cognitive

resources to organize the elements before being able to process them (see Section 2.5.2.2). Consequently, the learner is forced to use his limited cognitive resources to both organize and process information presented to him in the video. This can be straining on an L2 learner's working memory and cause a lower level of comprehension due to heavy cognitive load. This can be avoided with videos where input modes are coordinated, so that the need for organization is eliminated and the cognitive load is reduced.

Furthermore, multimedia designers usually try to adapt to individual learners' needs, who prefer either written or spoken text, and, in principle, will be able to choose their preferred input (Schnotz, 2014). However, scholars suggest that learners are unable to exclude written text when they are exposed to it, and, therefore, are forced to process both written and spoken text when presented concurrently (e.g. Schnotz, 2014). Thus, when watching videos that make use of presenting information in dual-mode, such as the 10 videos in the present study, L2 learners may experience a heavy cognitive load. They must process two modes in the same channel instead of placing their attention on their preferred mode, as is the intention of the video designers. In addition, as suggested by the redundancy principle (see section 2.5.2.3), when dually presented, either the spoken mode or written mode is redundant. This is because both modes contain the same information and one should be eliminated to ease the cognitive processing of information. This is supported in empirical research, as studies have shown that learners score lower on comprehension tests when visual input is combined with a dual-mode presentation of spoken and written text in combination with visual input (Kalyuga & Sweller, 2014).

4.2.2.1 Frequency of words in dual-mode

Based on the results presented in Table 4.4, words presented in dual-mode were analyzed for level of frequency using the AntWordProfiler software. The analysis followed the same method for collecting data as the first analysis, which examined the frequency levels of the total corpus (see Section 3.3.1). As Table 4.5 illustrates, 90.5% of the words in dual-mode were high-frequency words (between level 1,000-3,000), while no words were less frequent than the 7,000 level. Conclusively, this finding shows that the majority of words presented in dual-mode are high frequency words. Illustrated in Table 4.5 show, these words are mostly high-frequency words, which means that they may already be known by L2 learners with a vocabulary size of 3,000 word families or more. For these learners, words presented in dual-mode may be found redundant and cause unnecessary resources spent on processing. This can, in the worst case, cause cognitive overload and impair their comprehension of the videos.

For L2 learners with knowledge of less than 3,000 word families, however, words presented in dual-mode may help with lexical segmentation, as the words learners think they are hearing can be either confirmed or disproved in writing.

Finally, based on the previously mentioned theoretical framework, the findings from the present study imply that videos using text in dual-mode combined with pictures may hinder the comprehension of the videos if these modes are causing a split-attention effect, or if one mode is redundant. This may cause a heavy cognitive load for learners who must use limited cognitive resources to process information in the same verbal channel. Learners viewing these videos will need to divide their attention to accommodate all modes presented: the visual input, the written text, and the spoken text. If learners try to attend to all of the presented modes, their overall comprehension might be diminished due to their limited capacity of processing in each channel. Based on the empirical research mentioned in Section 2.5.2.3, the consensus amongst scholars is that the use of redundant text should be limited (e.g. Jamet & Le Bohec, 2007; Kalyuga, et. al, 2004).

4.2.3 General discussion

The aim of this study was to investigate the lexical demands placed on L2 learners when watching videos found on YouTube in the classroom. The study also examined how different types of input modes were used in these videos to better understand the relationship between the use of different input modes and L2 learners' listening comprehension. Thus, this section puts forward a general discussion of how lexical demands and different modes may have an impact on the L2 learners' comprehension of the videos.

The findings presented in the current thesis indicate that successful listening comprehension from videos depends on the close relationship between an L2 learner's word knowledge and how the video design provides the learner with contextual cues from visual input without exceeding his limited capacity in working memory.

The results presented in Table 4.1 show that the majority of words in the corpus belong to the first 1,000 word families (80.7%), with a decreasing number for each additional 1,000 word families as demonstrated by Zipf's law (see Section 2.4.2). These results lend cogency to the argument put forward by several scholars, e.g. Nation (2013), that high-frequency words should be learned before mid- and low-frequency words (see Section 2.4.2). If L2 learners do not have knowledge of the most frequently used words in the English language, they will have several difficulties when meeting authentic language situations, which are not adjusted to their level of language proficiency. This is also the case with the

videos analyzed in the current thesis. A majority of these videos may be too lexically demanding for L2 learners with a vocabulary size smaller than 3,000 word families, which support the argument that teachers need to be aware of the lexical demands they place on their learners when using videos such as these in the classroom.

Furthermore, the results from the second analysis show that all videos presented a combination of visual and verbal input, where the majority of videos presented words in spoken mode. Thus, it can be suggested that L2 learners will find support in visual input that will compensate for missing word knowledge, as proposed by the multimedia principle. When watching news videos, such as the videos in this thesis, reporters and interview object present in the video are visual elements which can aid L2 learners' comprehension. However, if the speaker is not shown, comprehension problems may occur at the speech perception stage when words are presented in spoken input mode. As research has suggested, this may impact the learner's listening comprehension negatively as it may make speech perception more difficult (see Section 2.3.1).

 Conclusively, if learners lack sufficient word knowledge, decoding of the aural input and, thus, comprehension may be hindered. Also, if there are multiple modes presented at the same time learners might experience a heavy cognitive load in addition to their difficulty of word knowledge. On the other hand, visual and verbal input modes, especially when combined, might help learners fill in the missing parts they might lack due to missing word knowledge. Conclusively, there is a close relationship between successful listening comprehension, word knowledge, and the use of different input modes in videos.

4.3 Chapter Summary

This chapter has put forward the findings from the analyses conducted in the present thesis. The findings were discussed in light of previously presented theory in order to answer the research questions guiding the study. As demonstrated, the results from the first analysis, which sought to answer the first research question, show that, although there are individual differences between the videos in the corpus, a vocabulary size of 3,000 word families, including proper nouns and marginal words are necessary to reach 95% coverage of the complete corpus. In order for L2 learners to reach 98% coverage of the total corpus, a vocabulary size of 5,000 word families will be sufficient.

The results from second analyses show that all videos present a combination of visual and verbal input modes, while half of the videos also made use of presenting words in dual-

mode. Within the theoretical framework surrounding the current thesis, these results suggest that when a combination of visual and verbal input modes is presented, comprehension may be enhanced as learners can use contextual cues to fill in their gaps of missing knowledge. However, when videos are using text in dual-mode, the combination may cause a decrease in the L2 learners' listening comprehension of the videos. This can be explained by either a split-attention effect or redundancy effect, which may lead to heavy cognitive load for the learners' limited working memory capacity. Additionally, in 11 videos non-essential written words appeared in the background. This may also impair L2 leaners' listening comprehension due to the imposed extraneous load learners might encounter. Furthermore, as the third analysis show, words presented in dual-mode were mostly high-frequency words which may already be known to many L2 learners. For these learners, this type of presentation may cause a redundancy effect and require learners to spend unnecessary cognitive resources on processing these words.

5. Conclusion

The aim of the present study has been to examine the lexical demands placed on L2 learners when presented with videos available on YouTube, and how these videos' use of different input modes may impact L2 learners' listening comprehension. Twenty videos were chosen from the BBC News YouTube channel and analyzed using Corpus Linguistics. The theoretical framework guiding the study was based on Mayer's (2001) cognitive theory of multimedia learning and Sweller's (1988) cognitive load theory, and the relevant principles discussed were the multimedia principle, the split-attention principle, and the redundancy principle.

The present study asked the following research questions:

1. What are the lexical demands placed on L2 learners when watching BBC News videos available on YouTube in the classroom?
2. How are different input modes presented in these videos?
   2a. How can the use of input modes in these videos impact listening comprehension?

The data collected in order to answer the research questions were gathered from a corpus comprised of transcripts from 20 BBC News videos available on YouTube, and quantitative methods were applied to the analyses. The study included an analysis of the lexical demands which were measured in terms of lexical coverage and vocabulary size. Additionally, it included an analysis of how different input modes were presented in the corpus, followed by a discussion of how the use of different input modes may or may not facilitate L2 learners' listening comprehension. The hope has been that the study may generate new knowledge about how lexically demanding BBC News videos are, and how the use of vocabulary and input modes in these videos impact L2 learners' listening comprehension. Moreover, with this thesis I sought to raise awareness of the lexical demands teachers may place on their learners when presenting videos found on YouTube in the language learning classroom.

In the present chapter, the key findings in relation to the research questions in the current thesis will be presented, followed by possible limitations, contributions, pedagogical implications, and suggestions for further research.

5.1 Key Findings

The following sections put forward the key findings in the current study. First, key findings from the first analysis will be presented. Second, the key findings from the second set of analyses will be highlighted.

5.1.1 Lexical coverage

This section puts forward the key findings related to the first research question. As presented in Chapter 2, word knowledge is an essential factor for successful listening comprehension (Rost, 2016). Thus, this thesis examined the frequency of vocabulary in a corpus of twenty BBC News videos. Moreover, to answer to the first research question, the analysis of the corpus showed that L2 learners will need to have knowledge of 3,000 word families for reaching 95% coverage, and 5,000 word families for reaching 98% coverage, including knowledge of proper nouns and marginal words. The results presented in this study are in compliance with previous studies examining movies and television programs (Webb & Rodgers, 2009a, 2009b), but suggests that these BBC News videos are slightly less demanding for reaching 98% coverage.

5.1.2 Input modes

This section puts forward the key findings related to the second set of research questions, which asked how different input modes were presented in these videos, and how this can impact L2 learners' listening comprehension. The data gathered in this thesis shows that the majority of words were presented as spoken input mode in the videos, in combination of visual input. Half of the videos presented words simultaneously in spoken and written mode (dual-mode), and a frequency analysis based on this finding shows that 90.48% of words presented in this dual-mode were high-frequency words. In relation to the three principles derived from the theories presented in this thesis, the present study discussed whether presenting words in two modes would enhance learning, as proposed by the multimedia principle, or if it might cause a split-attention effect or redundancy effect, and, consequently, cause a heavy cognitive load which will have a negative impact on L2 learners' listening comprehension.

5.2 Limitations

When conducting empirical research, there are bound to be limitations to the present study. In terms of the materials used for collecting data, the study is limited in the focus of only BBC

News videos, and the sample size is rather small. The focus on comprehension of these videos is also limited to include only listening. Thus, the study does not concern reading comprehension, although a lot of the input is presented as on-screen written text. A focus on viewing comprehension may have been more beneficial, as videos require more than just listening comprehension

The discussion of word knowledge concerned the unit of counting, receptive knowledge, and vocabulary size. The unit of counting in this study is the word family, which was defined in this study as the base word form plus its inflections and most common derivational variants. However, Nation (2006) suggests that most L2 learners will not have word families as inclusive as those of native speakers, and, consequently, the choice of word family level is of concern. The word families used in the word lists were based on Bauer and Nation (1993) level 6. For advanced learners, level 6 might be a suitable level, but that for less advanced learners, level 2 or 3 might be more appropriate. Using the appropriate word family level for Norwegian L2 learners in the lower secondary school would have added to the relevance of this study. Nevertheless, in order to be able to compare the present study's results to previous research, level 6 was chosen.

Some researchers argue for the use of lemmas (Schmitt, 2010), which contain all of the regularly inflected forms from the same stem and the same syntactic category. Consequently, lemmas distinguish between homonyms – which causes problems with the use of the AntWordProfiler software and other computer-based corpus analyzing programs that are incapable of distinguishing between homonyms. Word families, however, are different from lemmas in that they cross syntactic categories. Nation (2006) argues that the word family is the most sensible unit of counting for receptive purposes, but that if the word lists were made for productive purposes (for speaking and writing), the lemma would be the appropriate unit to use.

In this study, individual words were analyzed as separate entities (tokens). However, this study could have analyzed multiword units, as Schmitt (2010) suggests that formulaic language is pervasive, both in spoken and written discourse. Nevertheless, both kinds of learning are essential, and the focal point for this thesis has been on single-word units. Furthermore, according to Nation (2006), "[a]lthough transparent phrases need to be learned for productive purposes, for the receptive purposes of reading and listening they are not a major issue" (p. 66), which further supports the choice of limiting the scope to only concert single-word units.

For the secondary analysis, input modes were limited to four categories: spoken text, essential on-screen text, non-essential on-screen text, and on-screen text with aural narration. Thus, several input modes were excluded from the analysis, such as background music, background noise, still pictures, motion pictures, animations, and so on. The reason for this limitation was that these input modes would be difficult to measure through the use of the chosen data collection method, as well as the time constraints for the present study made such an analysis unachievable. In addition, the present study excluded speech rate and speed of input for the same reasons.

The focal point of this study has been L2 learners in the lower secondary school in Norway. However, as the videos selected for this study might also be used in the upper secondary school, this study may be relevant for teachers in the upper secondary level of education as well.

5.3 Contributions

It is my hope that the current thesis can contribute to a better understanding of the lexical demands teachers place on L2 learners when using videos from YouTube as teaching materials. I hope my findings will help raise teacher awareness of how their choice of videos can impact L2 learners' listening comprehension, and that it may encourage others to continue investigating the lexical demands of multimedia resources used in the EFL classroom. I also hope the study may contribute to greater insights into the processes involved when L2 learners listen to videos, and how missing vocabulary knowledge and understanding may be aided or hindered by visual elements, due to cognitive limitations of the working memory. I also hope that this study will contribute to raise a discussion of how much teachers can expect Norwegian L2 learners in the lower and upper secondary education to understand when watching to videos. There is, unfortunately, too little to be known about their expected receptive vocabulary size, as there seems to be a lack of empirical studies in this area. My study has commented on this need and hopes that such studies will emerge in the nearest future.

Furthermore, this thesis may also provide a better understanding of how computer software can be used to investigate lexical demands of different discourses, in hope that teachers may use such software as tools to analyze the level of their teaching materials. Knowledge of what vocabulary size certain discourse require, may provide teachers with an indication of what materials are appropriate for their learners. Finally, I hope a better

understanding of the relationship between vocabulary size and comprehension of videos, such as BBC News videos, has been provided with this study

5.4 Pedagogical Implications

Pedagogical implications in the present study involve deciding the appropriate vocabulary level for L2 learners in order to facilitate their listening comprehension of videos. YouTube videos can be powerful language learning resources in the EFL classroom, as they bring authentic language situation into the classroom that would otherwise not be present. The current study emphasizes that learning the most frequent 3,000 word families should be the minimal vocabulary goal for Norwegian L2 learners if these videos are to be used in the EFL classroom. However, if 98% coverage is the preferred coverage threshold, then understanding these BBC News videos may prove to be too difficult for many L2 learners and are only an appropriate source of L2 input for more advanced learners, unless additional support is given. Moreover, the results presented in this thesis suggest that there are individual differences between the videos, and that a vocabulary size of minimum 3,000 word families would be necessary for comprehension of the *total* corpus.

There are several ways teachers can minimize comprehension issues for learners with a smaller vocabulary size than required by these videos, as the findings in this thesis have indicated. For example, teachers can provide additional support, given in the form of glosses, pre-teaching vocabulary, slowing down the tempo of the video, or pause the video to discuss complicated sections before moving on. In their article, Watkins and Wilkins (2011) addresses various ways teachers can use the YouTube videos to teach English in an EFL classroom. Thus, together with the findings provided by this thesis, their article might be a useful supplement for teachers who want to know more about how they can use YouTube videos as teaching materials for teaching English to L2 learners. Furthermore, teachers should consider providing learners with captions in, as studies report that L2 learners find captions useful for speech decoding and meaning-making processes (Perez, Peters, & Desmet, 2014). However, not all videos have captions available.

Moreover, teachers need to consider the design of the videos because designers who have considered how the human mind works have designed videos that are more likely to lead to meaningful learning than those that are not (Mayer, 2014b). Consequently, when choosing videos as classroom materials, teachers should examine the multimedia design of the video to get a better understanding of how the design may impact the listening comprehension of their learners, as was thoroughly discussed in this thesis. It is important for teachers to remember

that these videos were not made for language learning, and so, the challenges L2 learners may face when presented with these videos should be carefully considered before being presented in the EFL classroom.

Conclusively, even with short videos, the lexical demands and the cognitive load videos yield on L2 learners' listening comprehension should be in focus for language teachers. Having awareness of how successful listening comprehension, L2 learners' vocabulary knowledge, and the use of input modes are closely related, will make the use of videos found on YouTube an effective language learning tool in the EFL classroom.

5.5 Further Research

The present study expands on earlier research in several ways. As far as I know, this is the first study to assess the lexical demands of BBC News videos available on YouTube. To the best of my knowledge, it is also the first study to examine the use of different input modes in videos such as these. Thus, the study has addressed the need for more extensive research on Norwegian L2 learners' listening comprehension abilities, especially in a multimedia context.

Follow-up studies similar to the one conducted here could be applied to a larger group of BBC videos to provide further insights into the lexical demands placed on learners when using similar videos as teaching materials in the EFL classroom. The videos selected for the purpose of the present study are only assumptions of what BBC videos might be used, or are used, in the EFL classroom in Norway. Studies investigating exclusively the YouTube videos Norwegian teachers use as classroom materials in an English language learning course may provide a better understanding of the lexical demands that are placed on Norwegian L2 learners in the classroom, and be of great benefit for teachers when choosing videos appropriate for the level of their learners.

The present study did not take into consideration the speech rate and speed of these videos. Rost (2016) emphasizes speech rate as one major factor of perceiving speech for L2 listeners, as L2 learners will have a hard time perceiving speech if the speech rate is too fast. It would have been interesting to analyze the speech rate in the videos examined in the present study to see how L2 listening comprehension might be affected by speed and speech rates. However, this was not within the scope of this study. Future studies should examine how the speech rate of these videos may impact L2 learners' listening comprehension, as this will further contribute to how videos found on YouTube will be comprehended by L2 learners.

Future research that examines the possibilities of vocabulary learning or incidental acquisition from watching these videos would be of great interest, as vocabulary knowledge is

an essential part of L2 learners' listening comprehension. A study of this type would add to the rationale of using YouTube videos in the language learning classroom.

This thesis did not assess how L2 learners might be impacted by the different uses of input modes but discussed what the potential outcomes could be for L2 learners' listening comprehension. A study testing how much L2 learners will be able to process when watching videos found on YouTube will contribute to a greater understanding of how much input L2 learners can be exposed to before exceeding their cognitive capacity. Furthermore, future studies assessing L2 learners' comprehension when listening to videos found on YouTube would greatly contribute to teachers' knowledge of the possible benefits of using YouTube in the EFL classroom.

As a last remark, it is important to mention that if other researchers want to replicate this thesis by using the same corpus of BBC News videos, their transcripts may not be similar to the transcripts making up this corpus, since transcripts are not available on the Internet. However, a set of principles were followed when transcribing the videos which can serve as a guideline for future studies (see Section 3.1.4.1).

References

Adolphs, S., & Schmitt, N. (2003). Lexical Coverage of Spoken Discourse. *Applied Linguistics, 24*(4), 425-438. doi:10.1093/applin/24.4.425

Anthony, L. (2014). AntWordProfiler (Version 1.4.1) [Computer Software]. Tokyo, Japan: Waseda University. Retrieved from http://www.laurenceanthony.net/software

Ayres, F., & Sweller, J. (2014). The Split-Attention Principle in Multimedia Learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2 ed.), 43-71. Cambridge: Cambridge University Press.

Baddeley, A. D. (1997). *Human memory: Theory and practice*: Psychology Press.

Baggett, P. (1989). Understanding Visual and Verbal Messages. *Advances in psychology, 58*, 101-124. doi:10.1016/S0166-4115(08)62150-0

Baltova, I. (1999). Multisensory language teaching in a multidimensional curriculum: The use of authentic bimodal video in core French. *Canadian modern language Review, 56*(1), 31-48.

Barcroft, J., & Wong, W. (2013). Input, Input Processing and Focus on Form. In J. Herschensohn & M. Young-Scholten (Eds.), *The Cambridge handbook of second language acquisition*. New York: Cambridge University Press.

Bauer, L., & Nation, P. (1993). Word Families. *International Journal of Lexicography, 6*(4), 253-279. doi:10.1093/ijl/6.4.253

BBC. (2019). About the BBC. Retrieved from https://web.archive.org/web/20100116202334/http://www.bbc.co.uk/info/purpose/what.shtml

Brett, P. (1997). A Comparative Study of the Effects of the Use of Multimedia on Listening Comprehension. *System, 25*(1), 39-53.

Brown, S. (2006). *Teaching listening.* New York: Cambridge University Press.

Brysbaert, M., Stevens, M., Mandera, P., & Keuleers, E. (2016). How many words do we know? Practical estimates of vocabulary size dependent on word definition, the degree of language input and the participant's age. *Frontiers in psychology, 7*.

Buck, G. (2001). *Assessing listening.* New York: Cambridge University Press.

Butcher, K. R. (2014). The Multimedia Principle. In R. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2 ed.), 174-205. New York: Cambridge University Press.

Clark, V., & Ivankova, N. (2015). *Mixed methods research: A guide to the field* (Vol. 3): Sage Publications.

Cobb, T. Compleat Web VP v.2 [Computer Program]. Retrieved from https://www.lextutor.ca/vp/comp/

Coxhead, A. (2000). A New Academic Word List. *TESOL quarterly, 34*(2), 213-238. doi:10.2307/3587951

Creswell, J. W. (2014). *Research design : qualitative, quantitative, and mixed methods approaches* (4th ed.). Los Angeles, California: SAGE.

Cross, J. (2009). Effects of listening strategy instruction on news videotext comprehension. *Language Teaching Research, 13*(2), 151-176.

Cross, J. (2011). Comprehending news videotexts: The influence of the visual content. *Language learning & technology, 15*(2), 44–68. doi:http://dx.doi.org/10125/44251

Cutler, A., & Carter, D. M. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech & Language, 2*(3-4), 133-142.

Danan, M. (1992). Reversed subtitling and dual coding theory - new directions for foreign-language instruction. *Lang. Learn., 42*(4), 497-527. doi:10.1111/j.1467-1770.1992.tb01042.x

Dang, T. N. Y., & Webb, S. (2014). The lexical profile of academic spoken English. *English for Specific Purposes, 33*, 66-76.

Dang, T. N. Y., & Webb, S. (2016). Evaluating lists of high-frequency words. *ITL-International Journal of Applied Linguistics, 167*(2), 132-158.

Dörnyei, Z. (2007). *Research methods in applied linguistics : quantitative, qualitative, and mixed methodologies*. Oxford: Oxford University Press.

Ellis, R. (1986). *Understanding second language acquisition*. Oxford: Oxford University Press.

Fleckenstein, L. M. (2015). *Vocabulary in 10th Grade - The receptive vocabulary size of 10th graders in Iceland.* (Bachelors BA), University of Iceland, Iceland.

Fletcher, J. D., & Tobias, S. (2005). The Multimedia Principle. In R. Mayer (Ed.), *The Cambridge handbook of multimedia learning*, 117-134. Cambridge: Cambridge University Press.

Gass, S. M., & Mackey, A. (2014). Input, Interaction, and Output in SLA. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition : an introduction* (2 ed.): Routledge Ltd - M.U.A.

Gilje, Ø., Ingulfsen, L., Dolonen, J. A., Furberg, A., Rasmussen, I., Kluge, A., . . . Skarpaas, K. G. (2016). *Med ARK&APP - Bruk av læremidler og ressurser for læring på tvers av arbeidsformer*. Oslo: Universitetet i Oslo

Goh, C. C. M. (2000). A cognitive perspective on language learners' listening comprehension problems. *System, 28*(1), 55-75.

Goulden, R., Nation, P., & Read, J. (1990). How Large Can a Receptive Vocabulary Be? *Applied Linguistics, 11*(4), 341-363. doi:10.1093/applin/11.4.341

Graddol, D. (2006). *English next* (Vol. 62): British Council London.

Gruba, P. (2006). Playing the videotext: A media literacy perspective on video-mediated L2 listening. *Language learning & technology, 10*(2), 77-92.

Guichon, N., & McLornan, S. (2008). The Effects of Multimodality on L2 Learners: Implications for CALL Resource Design. *System: An International Journal of Educational Technology and Applied Linguistics, 36*(1), 85-93. doi:10.1016/j.system.2007.11.005

Henriksen, B., & Danelund, L. (2015). Studies Of Danish L2 Learners' Vocabulary Knowledge And The Lexical Richness Of Their Written Production In English. *Lexical issues in L2 writing*.

Herschensohn, J., & Young-Scholten, M. (2013). *The Cambridge handbook of second language acquisition*. New York: Cambridge University Press.

Hu, M., & Nation, P. (2000). Unknown Vocabulary Density and Reading Comprehension. *Reading in a foreign language, 13*(1), 403-30.

Inceçay, V., & Koçoglu, Z. (2017). Investigating the Effects of Multimedia Input Modality on L2 Listening Skills of Turkish EFL Learners. *Education and Information Technologies, 22*(3), 901-916. doi:10.1007/s10639-016-9463-3

Jamet, E., & Le Bohec, O. (2007). The effect of redundant text in multimedia instruction. *Contemporary Educational Psychology, 32*(4), 588-598. doi:https://doi.org/10.1016/j.cedpsych.2006.07.001

Jones, L. (2001). Listening comprehension in multimedia learning: An extension of the generative theory of multimedia learning. In H. Preskill & J. Plass (Eds.): ProQuest Dissertations Publishing.

Kalyuga, S., Chandler, P., & Sweller, J. (1999). Managing split-attention and redundancy in multimedia instruction. *Applied Cognitive Psychology, 13*(4), 351-371.

Kalyuga, S., Chandler, P., & Sweller, J. (2004). When Redundant On-Screen Text in Multimedia Technical Instruction Can Interfere With Learning. *Human Factors: The*

Journal of Human Factors and Ergonomics Society, 46(3), 567-581. doi:10.1518/hfes.46.3.567.50405

Kennedy, G. (1998). *An introduction to corpus linguistics*. London: Longman.

Kozma, R. B. (1991). Learning with Media. *Review of Educational Research, 61*(2), 179-211. doi:10.3102/00346543061002179

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

Laufer, B., & Ravenhorst-Kalovski, G. C. (2010). Lexical Threshold Revisited: Lexical Text Coverage, Learners' Vocabulary Size and Reading Comprehension. *Reading in a Foreign Language, 22*(1), 15-30.

Laufer, B., & Sim, D. D. (1985). Measuring and explaining the reading threshold needed for English for academic purposes texts. *Foreign Language Annals, 18*(5), 405-411.

Lee, S.-P., Lee, S.-D., Liao, Y.-L., & Wang, A.-C. (2015). Effects of Audio-Visual Aids on Foreign Language Test Anxiety, Reading and Listening Comprehension, and Retention in EFL Learners. *Perceptual and Motor Skills, 120*(2), 576-590. doi:10.2466/24.PMS.120v14x2

Loewen, S., & Reinders, H. (2011). *Key concepts in second language acquisition*. Basingstoke: Palgrave Macmillan.

Lund, R. J. (1991). A Comparison of Second Language Listening and Reading Comprehension. *The Modern Language Journal, 75*(2), 196-204. doi:10.1111/j.1540-4781.1991.tb05350.x

Massaro, D. W. (1998). *Perceiving talking faces: From speech perception to a behavioral principle* (Vol. 1): Mit Press.

Mayer, R. E. (1997). Multimedia learning: Are we asking the right questions? *Educational Psychologist, 32*(1), 1-19.

Mayer, R. E. (2001). *Multimedia Learning*. Cambridge: Cambridge University Press.

Mayer, R. E. (2005). Cognitive Theory of Multimedia Learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning*, 31-48. Cambridge: Cambridge University Press.

Mayer, R. E. (2014a). *The Cambridge handbook of multimedia learning* (2 ed.). New York: Cambridge University Press.

Mayer, R. E. (2014b). Cognitive Theory of Multimedia Learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2 ed.), 43-71. Cambridge: Cambridge University Press.

Mayer, R. E. (2014c). Introduction to Multimedia Learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2 ed.). Cambridge: Cambridge University Press.

Mayer, R. E., & Sims, V. K. (1994). For whom is a picture worth a thousand words? Extensions of a dual-coding theory of multimedia learning. *Journal of Educational Psychology, 86*(3), 389-401. doi:10.1037/0022-0663.86.3.389

McEnery, T., & Hardie, A. (2012). *Corpus linguistics : method, theory and practice.* Cambridge: Cambridge University Press.

Milton, J. (2009). *Measuring second language vocabulary acquisition* (Vol. 45). Multilingual Matters.

Mitchell, R., Myles, F., & Marsden, E. (2013). *Second language learning theories* (3rd ed.). London: Routledge.

Moreno, R., & Mayer, R. E. (1999). Cognitive principles of multimedia learning: The role of modality and contiguity. *Journal of Educational Psychology, 91*(2).

Morgan, D. L. (2017). *Integrating qualitative and quantitative methods : a pragmatic approach.* USA: SAGE Publications.

Mousavi, S. Y., Low, R., & Sweller, J. (1995). Reducing cognitive load by mixing auditory and visual presentation modes. *Journal of educational psychology., 87*(2), 319-334.

Nation, P. (2001). Using small corpora to investigate learner needs. In M. Ghadessy, A. Henry, R. L. Roseberry, & J. Sinclair (Eds.), *Small Corpus Studies and ELT : Theory and practice*. Philadelphia: John Benjamins Publishing Company.

Nation, P. (2006). How Large a Vocabulary Is Needed for Reading and Listening? *The Canadian Modern Language Review = La Revue Canadienne Des Langues Vivantes., 63*(1), 59-82.

Nation, P. (2007). The Four Strands. *Innovation in Language Learning and Teaching, 1*(1), 2-13. doi:10.2167/illt039.0

Nation, P. (2013). *Learning vocabulary in another language* (2 ed.). Cambridge: Cambridge University Press.

Nation, P. (2014). How much input do you need to learn the most frequent 9,000 words? *Reading in a Foreign Language, 26*(2), 1-16.

Nation, P. (2017). The BNC/COCA Level 6 Word Family List (Version 1.0.0) [Data File]. Retrieved from http://www.victoria.ac.nz/lals/staff/paul-nation.aspx

Nation, P., & Waring, R. (1997). Vocabulary size, text coverage and word lists. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition, and pedagogy*, 6-19. Cambridge: Cambridge University Press.

Nation, P., & Webb, S. A. (2011). *Researching and Analyzing Vocabulary*: Heinle, Cengage Learning.

Nugent, G. C. (1982). Pictures, audio, and print: Symbolic representation and effect on learning. *ECTJ, 30*(3), 163-174.

Paivio, A. (1986). *Mental representations: A dual coding approach*. Oxford, England: Oxford University Press.

Patton, M. Q. (2002). *Qualitative research & evaluation methods* (3 ed.). Thousand Oaks, Calif: SAGE Publications.

Perez, M. M., Peters, E., & Desmet, P. (2014). Is less more? Effectiveness and perceived usefulness of keyword and full captioned video for L2 listening comprehension. *ReCALL, 26*(1), 21-43.

Plass, J. L., & Jones, L. C. (2005). Multimedia Learning in Second Language Acquisition. In *The Cambridge handbook of multimedia learning*, 467-488. Cambridge: Cambridge University Press.

Punch, K. (2003). *Survey research : the basics*. London: SAGE Publications.

Paas, F., & Sweller, J. (2014). Implications of Cognitive Load Theory for Multimedia Learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2 ed.), 43-71. Cambridge: Cambridge University Press.

Paas, F., van Gog, T., & Sweller, J. (2010). Cognitive Load Theory: New Conceptualizations, Specifications, and Integrated Research Perspectives. *Educational Psychology Review, 22*(2), 115-121. doi:10.1007/s10648-010-9133-8

Read, J. (1988). Measuring the vocabulary knowledge of second langauge learners. *RELC Journal, 19*(2), 12-25.

Reves, T., & Levine, A. (1988). The FL Receptive Skills: Same or Different? *System, 16*(3), 327-336.

Rindal, U., & Piercy, C. (2013). Being 'neutral'? English pronunciation among Norwegian learners. *World Englishes, 32*(2), 211-229. doi:10.1111/weng.12020

Rodgers, M. P. H., & Webb, S. (2011). Narrow Viewing: The Vocabulary in Related Television Programs. *TESOL Quarterly: A Journal for Teachers of English to Speakers of Other Languages and of Standard English as a Second Dialect, 45*(4), 689-717. doi:10.5054/tq.2011.268062

Rost, M. (2002). *Listening tasks and language acquisition.* Paper presented at the Memorias del Congreso JALT 2002.

Rost, M. (2005). L2 Listening. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning*, 503-527. Mahwah, N.J: Lawrence Erlbaum.

Rost, M., & Ross, S. (1991). Learner use of strategies in interaction: Typology and teachability. *Language Learning, 41*(2), 235-268.

Rubin, J. (1995). The contribution of video to the development of competence in listening. *A guide for the teaching of second language listening*, 151-165.

Schmitt, N. (2000). *Vocabulary in language teaching*. Cambridge: Cambridge University Press.

Schmitt, N. (2010). *Researching vocabulary : a vocabulary research manual*. Basingstoke: Palgrave Macmillan.

Schmitt, N., Jiang, X., & Grabe, W. (2011). The Percentage of Words Known in a Text and Reading Comprehension. *The Modern Language Journal, 95*(1), 26-43.

Schmitt, N., & Schmitt, D. (2014). A Reassessment of Frequency and Vocabulary Size in L2 Vocabulary Teaching. *Language Teaching: Surveys and Studies, 47*(4), 484-503. doi:10.1017/S0261444812000018

Siddell, F. (2011). Sounds Comprehensible: Using Media for Listening Comprehension in the Language Classroom. In W. M. Chan, K. N. Chin, M. Nagami, & T. Suthiwan (Eds.), *Media in Foreign Language Teaching and Learning*: Retrieved from https://ebookcentral.proquest.com.

Stæhr, L. S. (2008). Vocabulary size and the skills of listening, reading and writing. *The Language Learning Journal, 36*(2), 139-152. doi:10.1080/09571730802389975

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive science, 12*(2), 257-285.

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive Load Theory*: Springer New York

Tegge, F. (2017). The Lexical Coverage of Popular Songs in English Language Teaching. *System: An International Journal of Educational Technology and Applied Linguistics, 67*, 87-98. doi:10.1016/j.system.2017.04.016

The Norwegian Directorate for Education and Training (2013). *English subject curriculum (ENG1-03)*. Retrieved from https://www.udir.no/kl06/ENG1-03?lplang=http://data.udir.no/kl06/eng.

Tognini-Bonelli, E. (2001). *Corpus Linguistics at Work*. Amsterdam: John Benjamins Publishing Company.

Tognini-Bonelli, E. (2010). *Theoretical overview of the evolution of corpus linguistics*. London: Routledge.

Tschirner, E. (2011). Video Clips, Input Processing And Language Learning. In W. M. Chan, K. N. Chin, M. Nagami, & T. Suthiwan (Eds.), *Media in foreign language teaching and learning* (Vol. 5). Boston: De Gruyter Mouton.

van Zeeland, H., & Schmitt, N. (2012). Lexical Coverage in L1 and L2 Listening Comprehension: The Same or Different from Reading Comprehension? *Applied Linguistics, 34*(4), 457-479. doi:10.1093/applin/ams074

Vandergrift, L. (2006). Second Language Listening: Listening Ability or Language Proficiency? *The modern language journal., 90*(1), 6-18.

Vandergrift, L. (2007). Recent Developments in Second and Foreign Language Listening Comprehension Research. *Language teaching., 40*(3), 191-210.

Vandergrift, L., & Goh, C. (2009). Teaching and Testing Listening Comprehension. In M. H. Long & C. Doughty (Eds.), *The handbook of language teaching*. Malden, Mass: Wiley-Blackwell.

VanPatten, B. (2014). Input Processing in Adult Second Language Acquisition. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition : an introduction* (2 ed.): Routledge.

VanPatten, B., & Williams, J. (2007). *Theories in second language acquisition : an introduction*. Mahwah, N.J: Lawrence Erlbaum.

Watkins, J., & Wilkins, M. (2011). Using YouTube in the EFL Classroom. *Language Education in Asia, 2*(1), 113-119. doi:10.5746/LEiA/11/V2/I1/A09/Watkins_Wilkins

Webb, S. (2008). Receptive and productive vocabulary sizes of L2 learners. *Studies in Second Language Acquisition, 30*(1), 79-95.

Webb, S., & Nation, I. S. P. (2017). *How vocabulary is learned*. Oxford: Oxford University Press.

Webb, S., & Rodgers, M. P. H. (2009a). The Lexical Coverage of Movies. *Applied Linguistics, 30*(3), 407-427. doi:10.1093/applin/amp010

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

West, M. (1953). *A General Service List of English Words* London: Longman, Green & Co.

Wetzel, C. D., Radtke, P. H., & Stern, H. W. (1993). *Review of the effectiveness of video media in instruction. No. NPRDC-TR-93-4*. Navy Personnel Research and Development Center, San Diego, California

Yang, J. C., & Chang, P. (2014). Captions and reduced forms instruction: The impact on EFL students' listening comprehension. *ReCALL, 26*(1), 44-61. doi:10.1017/S0958344013000219

Zipf, G. K. (1949). *Human behavior and the principle of least effort*. Cambridge: Addison-Wesley Press.

Appendix I. Survey Questions

Formålet med denne undersøkelsen er å finne ut av hvilke YouTube videoer og YouTube kanaler som blir brukt i engelskundervisningen på ungdomsskoler i Norge. Denne undersøkelsen ønsker altså å finne eksempler på YouTube-videoer brukt til undervisningsformål i <u>engelsk</u> og det er derfor viktig at du bruker god tid på å finne eksempler du selv bruker i undervisningen – hvis du bruker YouTube i undervisningen din. Bruk gjerne tid på å finne tilbake til YouTube-videoer du har brukt eller planlegger å bruke, slik at du kan skrive dem ned i undersøkelsen. Alle YouTube-videoer er av interesse, så lenge de har blitt/skal bli brukt i klasserommet - uansett lengde, nivå, innhold osv.

NB. Undersøkelsen gjelder bare YouTube-videoer brukt til undervisning av <u>engelsk</u>. Undersøkelsen vil ta deg ca. 5-10 minutter å gjennomføre.

Personvern: Det er frivillig å delta i prosjektet. Hvis du velger å delta, kan du når som helst trekke samtykke tilbake uten å oppgi noen grunn. Det vil ikke ha noen negative konsekvenser for deg hvis du ikke vil delta eller senere velger å trekke deg. Hvis du velger å delta i prosjektet, innebærer det at du fyller ut et spørreskjema. Dine svar fra spørreskjemaet blir registrert elektronisk. Spørreskjemaet spør ikke om personlige opplysninger som kan gjøre deg gjenkjennelig for andre. Spørreskjemaet er laget gjennom SurveyXact som skal sikre ditt personvern og holde deg anonym. SurveyXact sikrer anonymitet ved å sperre all tilgang til svarene du har gitt, samt respondentnøklene som gir tilgang til svarene, for brukeren av programmet. Alle data vil bli slettet etter prosjektslutt (juni, 2019).

Vi behandler opplysninger om deg basert på ditt samtykke.

På oppdrag fra UiB/Institutt for fremmedspråk har NSD – Norsk senter for forskningsdata AS - vurdert at behandlingen av personopplysninger i dette prosjektet er i samsvar med personvernregelverket.

Jeg bekrefter med dette at jeg samtykker til å delta i undersøkelsen.

(1) ❑ Ja

(2) ❑ Nei

En **YouTube-video** er alle typer videoer som blir brukt gjennom nettsiden www.youtube.com - uavhengig av lengde.

Bruker du YouTube i engelskundervisningen din?

(1) ❑ Ja

(2) ❑ Nei

Hvor ofte bruker du i gjennomsnitt YouTube-videoer til undervisningsformål?

(1) ❑ Sjeldnere enn én gang i måneden

(2) ❑ én gang i måneden

(3) ❑ to til tre ganger i måneden

(4) ❑ én gang i uken

(5) ❑ to til tre ganger i uken

(6) ❑ Hver dag

En **YouTube-kanal** defineres her som et sted på YouTube der en bruker kan legge ut flere videoer gjennom et brukernavn og som man kan abonnere på. En bruker trenger ikke være en privatperson men kan også være en bedrift/organisasjon som legger ut flere videoer gjennom kanalen sin

Bruker du noen av disse YouTube-kanalene til undervisning av engelsk? (Velg så mange du vil)

(1) ❑ Ted-Ed

(2) ❑ CrashCourse

(3) ❑ Vox

(4) ❑ Khan Academy

(5) ❑ Nasjonal digital læringsarena (NDLA)

(6) ❑ BBC

(7) ❑ Annet, vennligst spesifiser _____

(8) ❑ Ingen av de nevnte

Hvordan velger du YouTube-videoer til bruk i engelskundervisningen? Beskriv din fremgangsmåte

_____

_____

_____

_____

_____

_____

Til hvilke(t) formål bruker du vanligvis YouTube-videoer i undervisningen? (velg én eller flere)

(1) ❑ Til introduksjon av nytt tema

(2) ❑ Som underholdning

(3) ❑ For å oppsummere

(4) ❑ For å repetere

(5) ❑ For å gjøre oppgaver tilhørende videoen

(6) ❑ Annet _____

Nevn én eller flere YouTube-kanaler du bruker i engelskundervisningen

_____

_____

_____

_____

_____

_____

Nevn én eller flere YouTube-videoer du bruker i undervisningen din (tittel på video)

_____

_____

_____

_____

_____

_____

Bruker du kortere videoklipp (under 30 minutter) fra andre nettsteder/plattformer enn www.Youtube.com?

(1) ❑ Ja
(2) ❑ Nei

Hvilke klipp viser du og fra hvilket nettsted/plattform er de fra?

_____

_____

_____

_____

_____

_____

Hvorfor bruker du ikke YouTube i engelskundervisningen din?

(1) ❑ Jeg har ikke utstyret som trengs for å vise YouTube-videoer i klasserommet

(2) ❑ Jeg synes det er vanskelig å finne videoer som passer til undervisningen min

(3) ❑ Jeg vet ikke hva YouTube er

(4) ❑ Jeg bruker ikke digitale hjelpemidler i min undervisning

(5) ❑ Jeg tror ikke elevene mine forstår videoene som blir vist

(6) ❑ Jeg viser bare lengre (over 60 min.) filmer

(7) ❑ Annet _____

Tusen takk for at du tok deg tid til å svare på denne undersøkelsen!

Appendix II. Overview of Videos, Titles, and Time

Video	Title and	Time
1	Five things to know about US child migrant separations - BBC News https://www.youtube.com/watch?v=X2s_cbrfvS4&list= PLSPIPJfn1CMawH74M1-sIBt9gfEL1LkT_&index=1	01:50
2	Inside the mind of white America - BBC News https://www.youtube.com/watch?v=dykgtGv8ozs&list= PLSPIPJfn1CMawH74M1-sIBt9gfEL1LkT_&index=2	09:15
3	Do you want to be the President of the US? - BBC News https://www.youtube.com/watch?v=D_wMD5tohgA&list= PLSPIPJfn1CMawH74M1-sIBt9gfEL1LkT_&index=18	02:09
4	Baltimore: 'This is what poverty in the US looks like' - BBC News https://www.youtube.com/watch?v=tCgqIN-6A20&index=4&list= PLSPIPJfn1CMawH74M1-sIBt9gfEL1LkT_	02:24
5	Where do you fit on the US political spectrum? - BBC News https://www.youtube.com/watch?v=gZAYMWSBXCg&list= PLSPIPJfn1CMawH74M1-sIBt9gfEL1LkT_&index=3	02:01
6	Why is US in love with football? Sorry, soccer - BBC News https://www.youtube.com/watch?v=uad9QusbhPc&list= PLSPIPJfn1CMawH74M1-sIBt9gfEL1LkT_&index=5	02:37
7	The US war that never officially ended - BBC News https://www.youtube.com/watch?v=LLyhO42aUvs&list= PLSPIPJfn1CMawH74M1-sIBt9gfEL1LkT_&index=6	02:17
8	US Supreme Court backs gay marriage - BBC News https://www.youtube.com/watch?v=kj0SFX4HSag&list= PLSPIPJfn1CMawH74M1-sIBt9gfEL1LkT_&index=7	00:50
9	The winners and losers in US tax bill - BBC News https://www.youtube.com/watch?v=oyVn53-clR4&list= PLSPIPJfn1CMawH74M1-sIBt9gfEL1LkT_&index=8	02:13
10	Is now the best time to be young in the US? - BBC News https://www.youtube.com/watch?v=UJz5c6aaatU&list= PLSPIPJfn1CMawH74M1-sIBt9gfEL1LkT_&index=9	03:04

11	US gun laws: 'Obama has no right to do this' - BBC News https://www.youtube.com/watch?v=oeLV4sAUqV0&index=10&list= PLSPIPJfn1CMawH74M1-sIBt9gfEL1LkT_	01:44
12	US students head to Germany for free degrees - BBC News https://www.youtube.com/watch?v=oCtqHwjCPOQ&list= PLSPIPJfn1CMawH74M1-sIBt9gfEL1LkT_&index=11	03:06
13	US soccer: Will Americans ever care about football? - BBC News https://www.youtube.com/watch?v=VyXgE6H6tms&list= PLSPIPJfn1CMawH74M1-sIBt9gfEL1LkT_&index=12	04:16
14	Trump and the facts about the migrant caravan - BBC News https://www.youtube.com/watch?v=ETUqj2Fi9ZA&list= PLSPIPJfn1CMawH74M1-sIBt9gfEL1LkT_&index=13	02:44
15	Mexican building US wall: It's a job - BBC News https://www.youtube.com/watch?v=mucl873ukjU&list= PLSPIPJfn1CMawH74M1-sIBt9gfEL1LkT_&index=14	02:03
16	New security measures to prevent shootings in US schools - BBC News https://www.youtube.com/watch?v=ZG8Dc-dFDnI&list= PLSPIPJfn1CMawH74M1-sIBt9gfEL1LkT_&index=15	03:29
17	The moment after US Supreme court same-sex marriage decision - BBC News https://www.youtube.com/watch?v=jkHz-EaWxCY&list= PLSPIPJfn1CMawH74M1-sIBt9gfEL1LkT_&index=16	01:51
18	Who are the children crossing the US border? - BBC News https://www.youtube.com/watch?v=9XE1tCF5d5w&list= PLSPIPJfn1CMawH74M1-sIBt9gfEL1LkT_&index=17	02:06
19	US election 2016: Life on the US-Mexico border - BBC News https://www.youtube.com/watch?v=sqNYvXcSW5o&list= PLSPIPJfn1CMawH74M1-sIBt9gfEL1LkT_&index=19	02:26
20	Trump v Obama: Battle of the presidents - BBC News https://www.youtube.com/watch?v=6XQjsH5osUA&list= PLSPIPJfn1CMawH74M1-sIBt9gfEL1LkT_&index=20	02:33