RESEARCH ARTICLE

Lexical coverage in L1 and L2 viewing comprehension

Marion Durbahn[1,2] (iD), Michael Rodgers[3] (iD), Marijana Macis[4] (iD) and Elke Peters[1] (iD)

[1]KU Leuven, Research Group Language, Education, & Society, Pontificia Universidad Católica de Chile, Campus San Joaquín; [2]Pontificia Universidad Católica de Chile, Campus San Joaquin; [3]Carleton University, Ottawa, Canada and [4]Manchester Metropolitan University, Manchester, United Kingdom
Corresponding author: Marion Durbahn; Email: Marion.durbahn@kuleuven.be

Abstract

This study aimed to investigate the relationship between lexical coverage and TV viewing comprehension. Previous studies have indicated that 95% to 98% of lexical coverage may be needed for reading comprehension (Hu & Nation, 2000). To understand informal listening passages, lower coverage figures (95%-90%) may suffice. However, no study has researched the lexical coverage needed to understand audiovisual texts. We adopted a counter-balanced within-participants design, in which 5%, 10%, or 20% of the words in four 2-min documentaries were replaced with nonwords. Native and non-native speakers of English participated in this study. Results showed that comprehension scores decreased as lexical coverage decreased; comprehension at 100% coverage was significantly higher than 90% and 80% in the two groups; and optimal adequate comprehension is achieved with an optimal lexical coverage of 95%, whereas minimal adequate comprehension is reached with a minimal lexical coverage of 80%.

Keywords: Lexical coverage; viewing comprehension; nonwords

Lexical coverage generally refers to the percentage of running words known in a piece of text (Nation, 2001, 2006; van Zeeland & Schmitt, 2013, Webb, 2021). Studies on lexical coverage can be traced back to the 1980s with the advent of technological advancements for text analysis (Nurmukhamedov & Webb, 2019) and different methodologies have been used to calculate it. A common methodology for studying the relationship between lexical coverage and comprehension is to manipulate a text by replacing words with nonwords (Giordano, 2021; Hu & Nation, 2000; van Zeeland & Schmitt, 2013), whereas other studies have tested learners on a sample of words occurring in the input to determine coverage levels (Schmitt et al., 2011; Durbahn et al., 2020). Another methodology used in lexical coverage studies is corpus-based vocabulary profiling to

determine what lexical-level learners would need to understand that input (Nation, 2006; Webb, 2021). Finally, lexical coverage has also been gauged by relating learners' results on a frequency-based vocabulary test, such as the Vocabulary Levels Test, with a certain text's lexical profile to indicate the level learners would need to understand a certain text (Laufer & Ravenhorst-Kalovski, 2010; Noreillie, et al., 2018; Stæhr, 2009).

Research into lexical coverage has shown the importance of vocabulary knowledge for comprehension (Webb, 2021), and findings have helped set vocabulary targets for second-language (L2) learners to understand different input modes. There is considerably less research on viewing comprehension and lexical coverage than other modes, such as reading or listening. However, findings from reading and listening research are not sufficient to explain the effect of lexical coverage on viewing comprehension because of the presence of imagery. According to the Cognitive Theory of Multimedia Learning (Mayer, 2009), comprehension is enhanced when verbal and visual information are presented simultaneously. Because the imagery in audiovisual input may support the understanding of unknown words from context, it remains unclear how lexical coverage contributes to viewing comprehension. The present study aims to fill the above-mentioned gaps by investigating the effects of the degree of lexical coverage on viewing comprehension. A study like this has theoretical importance because it stresses the need to treat viewing as a construct by itself and not as a form of listening. Additionally, it is necessary to set input-specific goals in language programs.

Background

Studies on Lexical Coverage and L2 Reading Comprehension

The Lexical Quality Hypothesis outlines word knowledge as central to reading comprehension (Perfetti & Stafura, 2014). Although the relationship between lexical coverage and L2 reading comprehension has been widely explored (Laufer, 1989, 1992; Laufer & Sim, 1985; Qian 1999; Hu and Nation 2000), four studies are particularly relevant to the present investigation because of their findings and methodology.

Hu and Nation (2000) examined the percentage of lexical coverage needed for unassisted L2 reading for pleasure. They created four versions of a fictional text by replacing words with nonwords:

1. The 100% coverage version containing no nonwords,
2. The 95% coverage version with 5% nonwords,
3. The 90% coverage version with 10% nonwords, and
4. The 80% coverage version with 20% nonwords.

The authors ensured that all the remaining words belonged to the 2000 most frequent word families to avoid vocabulary difficulties. To assess comprehension, a multiple-choice and a written recall test were used. Results showed the following:

- The density of unknown words affected comprehension;
- No learners gained adequate comprehension at 80% lexical coverage; and
- Although it was possible for a few learners to gain adequate comprehension at a level as low as 95% and even 90%, most learners did not.

Due to the limited number of learners obtaining adequate comprehension at 90% to 95%, Hu and Nation (2000) asserted that, although they did not specifically investigate the 98%

coverage level, it would be the most reliable lexical coverage for most learners to attain unassisted comprehension of a fiction text. Therefore, it is the coverage suggested by the authors. It should be noted that Hu and Nation defined adequate comprehension as obtaining 12 of 14 correct answers on the multiple-choice test and 70 of 124 on the written recall test, which was the most recurrent score on the comprehension tests (mode).

Research by Laufer and Ravenhorst-Kalovski (2010) suggested a coverage level of 95% to 98% for comprehension of written texts. In their study, the authors explored the relationship between adequate reading comprehension, lexical coverage, and vocabulary knowledge. Reading comprehension was assessed by means of the English part of a standardized university entrance test. Unlike Hu and Nation (2000), Laufer and Ravenhorst-Kalovski (2010) did not replace words with nonwords to manipulate the degree of lexical coverage. Instead, they estimated lexical coverage by comparing participants' results on the revised version of Nation's (1983) vocabulary levels text (VLT; Schmitt, Schmitt, & Clapham, 2001) with the text vocabulary profiling. That is to say, if participants were familiar with the 3000 most frequent word families, they were expected to have 90% lexical coverage. Results showed that the higher the lexical coverage, the better the reading scores. The authors, then, suggest an optimal threshold reached at 98% lexical coverage, which corresponds to knowledge of the 6000 to 8000 most frequent word families, and a minimal threshold obtained at 95% lexical coverage with knowledge of the 4000- to 5000-word families. Adequate comprehension was defined as obtaining 134 of 150 on the standardized entrance test.

Schmitt et al. (2011) argued that the findings from Hu and Nation (2000) clearly showed that comprehension increases when vocabulary knowledge increases. The authors explored the relationship between the percentage of known words in a text and the degree of reading comprehension. Instead of manipulating a text (Hu & Nation, 2000) or using a VLT (Laufer & Ravenhorst-Kalovski, 2010) to determine lexical coverage, the authors tested participants on a large sample of the words occurring in two academic texts. To that end, they developed a checklist vocabulary test. A learner's lexical coverage was determined based on their score on the checklist test. That is to say, scores were split into 1% coverage bands (viz. 99%, 98%, 97%, etc.). Comprehension was measured through a multiple-choice test and a graphic organizer. Participants' mean comprehension scores were calculated for each coverage percentage. Results of the study indicated that there was a positive, medium-sized linear correlation ($r = .41$), between lexical coverage and reading comprehension. The lexical coverage needed would depend on the degree of comprehension required. If 60% was considered adequate, then 95% lexical coverage should be targeted, whereas when 70% is necessary, then 98% to 99% is required. Schmitt et al. (2011) suggest that 98% coverage is a reasonable target for reading academic texts.

The most recent research on the relationship between lexical coverage and reading comprehension is the one conducted by Song and Reynolds (2022). In their study, the authors examined the effect of two language learning variables (i.e., lexical coverage and topic familiarity) on L2 comprehension of expository texts by controlling for L2 reading ability and vocabulary size. Like Hu and Nation (2000), the authors manipulated two texts by replacing words with nonwords to create six levels of lexical coverage: 100%, 99%, 98%, 97, and 96%. Comprehension was measured through two reading comprehension tests, one for a familiar topic and one for an unfamiliar topic. Results revealed a nonsignificant interaction between topic familiarity and lexical coverage on comprehension scores. Although topic familiarity exhibited a significant main effect, lexical coverage did not. This suggests that, with controlled L2 vocabulary size and reading ability, topic familiarity has a larger effect on the expository comprehension of L2

learners than lexical coverage. These findings shed more light on the lexical coverage–comprehension relationship, adding to the results from previous studies that primarily focused on the role of lexical coverage in L2 reading comprehension. Regardless of lexical coverage, topic familiarity enhanced expository reading comprehension. This underscores the vital role of subject matter in text comprehension, enabling readers to filter out irrelevant information and potentially facilitating the inference of unknown words' meanings. While acknowledging the importance of lexical coverage, the study supports the claim that it is just one of the many factors influencing L2 reading comprehension. These findings align with recent research that suggests that the relationship may be more complex than initially thought.

Two recent meta-analyses about the relationship between vocabulary knowledge and reading comprehension have found a correlation ranging from .57 (unattenuated correlation) to .67 (corrected correlation for attenuation) in Zhang and Zhang (2020) and .79 in Jeon and Yamashita (2014). Results of the meta-analyses indicate that vocabulary knowledge accounts for more than a 31% variance in L2 reading comprehension (Zhang & Zhang, 2020).

In summary, previous research has agreed that 95% to 98% lexical coverage is necessary to comprehend a text adequately. Although more recent research has found that topic familiarity may be a better predictor of text comprehension, the two meta-analyses have shown that vocabulary knowledge can still predict at least 31% of the variance in L2 reading.

Studies on Lexical Coverage and L2 Listening Comprehension

Although more studies have been conducted on the relationship between lexical coverage and reading than on listening, the number of studies focusing on listening is on the rise. Two approaches to the study of lexical coverage and listening are discussed in this section: studies using VLTs matched with text profiling to estimate whether learners have the lexical coverage necessary to understand the input (Noreillie et al., 2018; Stæhr, 2009) and studies manipulating the lexical coverage by replacing words with nonwords (Giordano, 2020; van Zeeland & Schmitt, 2013).

Stæhr (2009) examined the correlation between vocabulary size and depth and adequate listening comprehension at an advanced proficiency level. Danish English as a foreign language participants took the second version of Schmitt et al.'s (2001) VLT, a vocabulary depth test (an adapted version of the Word Associates Test), and a listening comprehension test (Cambridge Certificate of Proficiency in English). To estimate participants' lexical coverage, the listening passages were submitted to a lexical frequency analysis, matched to the participants' results on the VLT, and related to the listening comprehension scores. Therefore, if learners were familiar with the 3K level in the VLT, they were assumed to know 94% of the words in the text and had a mean comprehension score of 59.1%, whereas knowledge of the 5K would correspond to the understanding of 98% of the text and lead to a mean comprehension score of 72.9%. Adequate comprehension was operationalized as obtaining at least 70% on the comprehension test. Results indicated that vocabulary size and depth were significantly correlated with listening comprehension (.70 and .65, respectively), with vocabulary size a major predictor of listening comprehension. Furthermore, the author explains that a lexical coverage of 98% and a vocabulary target of 5000 words would largely facilitate listening comprehension and would allow learners to reach a mean comprehension score of 73% on an advanced listening test.

Whereas Staehr focused on an advanced level, Noreillie et al. (2018) looked at the relationship between lexical coverage and listening comprehension at an intermediate level. In their adapted design, they conducted two studies. In the first, learners took the Cambridge Preliminary English Test (PET) and a self-designed vocabulary test (VocabLab test; Peters Velghe, & Van Rompaey, 2015). In the second study, participants took a French vocabulary test and the Diplôme d'études en langue française (DELF), a listening comprehension test for French. Findings indicated that there was a positive, strong correlation, r_s = .63, between lexical coverage and listening comprehension at an intermediate level. Additionally, they found that a lower coverage of approximately 90% may be sufficient to achieve adequate comprehension. Adequate comprehension was operationalized as obtaining 70% on the PET and 62% on the DELF test. One limitation of the study was that written VLTs are designed to measure written word form, which might not directly reflect the participants' knowledge of the spoken form of words.

Van Zeeland and Schmitt (2013) used a different approach to studying lexical coverage and listening. They drew on Hu and Nation's (2000) method and used nonwords to create texts with different degrees of lexical coverage. Furthermore, they included first-language (L1) participants in addition to L2 participants to investigate how lexical coverage affected L1 learners' listening comprehension skills. The authors manipulated four spoken, informal narrative passages by replacing nonwords to create four different coverage levels: 100%, 98%, 95%, and 90%. Results indicated that there was no statistical difference between comprehension scores at 95% and 90% coverage levels as both led L1 and L2 speakers to adequately comprehend the spoken texts. However, because L2 speakers showed greater variation in comprehension scores at the 90% level than at the 95% level, the authors concluded that 95% coverage may be needed for comprehension of informal texts. If detailed comprehension is needed, then 98% coverage may be the optimal coverage level. Given that the texts used for this study were everyday storytelling, the results may not be applicable to other genres.

One of the latest studies on the relationship between lexical coverage and listening comprehension was conducted by Giordano (2021). In his quasi-experimental research, the author inserted nonwords in five dialogs to create different levels of lexical coverage (viz. 98%, 95%, 90%, 85%, and 83%) at an intermediate level of English. The author also explored other variables that might affect comprehension, such as speech rate, topic familiarity, and discourse structure. Adequate comprehension was defined as obtaining 7 of 10 correct answers. Results supported previous findings on listening with learners having adequate comprehension at 90% lexical coverage. However, learners rated the 90% lexical coverage dialogue as easier than the other dialogs, which may explain why that lexical coverage level had the higher comprehension scores.

A recent meta-analysis found a correlation ranging from .56 (unattenuated) to .67 (corrected for attenuation; Zhang & Zhang, 2020) between vocabulary knowledge and listening comprehension. The reason for the slightly lower correlation between listening and reading comprehension (Jeon & Yamashita, 2014; Zhang & Zhang, 2020) may lie in the differences between spoken and written texts. Listening comprehension entails phonological knowledge and strategic competence to process the incoming stream of speech quickly and automatically (Stæhr, 2008). Additionally, listening contains nonverbal aspects, such as rhythm, emphasis, tone, etc., that may facilitate comprehension (Durbahn et al., 2020). Therefore, learners may not only rely on vocabulary knowledge but also on the other aspects of spoken language to understand listening texts.

Studies on Lexical Coverage and L2 Viewing Comprehension

The relationship between lexical coverage and viewing comprehension remains under-explored. Given that the imagery present in videos may help understand the meaning of the input, the relationship between lexical coverage and viewing comprehension may be different than with reading or listening. Mayer's (2009) Cognitive Theory of Multimedia Learning postulates that "People learn better from words and pictures than from words alone" (p. 1). Multimedia learning, also referred to as "dual-mode, dual-format, dual-code, or dual-channel learning" (p. 5), attempts to explain how the human mind processes information. Three assumptions underlie the theory (Mayer, 2014): dual channels, limited capacity, and active processing.

The dual-channel assumption, rooted in Paivio's (1986) Dual-Coding Theory, explains that humans process information through two separate channels (viz. auditory and visual channels). Although separate, learners may be able to convert the information represented in one channel and process it in the other channel. When information is presented in a more engaging and multisensory manner, it can enhance learning and, therefore, comprehension.

The Limited Capacity assumption is that humans can effectively process a finite amount of information in each channel, such as a few words (auditory channel) or a few images (visual channel). Well-constructed materials can enhance comprehension by reducing cognitive load, avoiding overload, focusing attention, and prioritizing essential information. This ensures that learners can assimilate and internalize the information effectively, leading to improved comprehension of the presented input.

The Active Processing Assumption is that humans take an active role in mentally processing information to form a cohesive representation of their experiences. This involves actively selecting pertinent incoming data, organizing it into a coherent cognitive framework, and integrating it with existing knowledge. Mayer's Cognitive Theory of Multimedia Learning can be closely related to viewing comprehension since watching a video would represent a multimedia experience where words and images are presented in close proximity. That proximity was found to be more recurrent in the case of documentaries (Rodgers, 2018).

In light of the definitions provided above, Mayer (2014) proposes that multimedia instruction should be designed aiming at taking full advantage of how the human mind works (i.e., assuming that humans have two information processing systems; that humans can process a limited amount of information in each channel, but that the information of each channel can be combined; and that humans actively engage in cognitive processes to make sense of multimedia presentations). When material is presented using only one channel, as in reading only and listening only, the potential contribution of the other channel is ignored (Mayer, 2009). Although Mayer's theory is primarily focused on learning, enhanced learning processes often result in better comprehension of the input.

Less research has been carried out on viewing comprehension. Two corpus studies have been conducted in which the authors performed a lexical frequency analysis of a large corpus of TV programs and movie scripts (Webb & Rodgers, 2009a, 2009b). The aim was to determine the vocabulary demands needed to reach 95% and 98% coverage of the words in TV shows and movies. Those figures were borrowed from the percentages suggested in reading research to achieve minimal and adequate comprehension (Laufer & Ravenhorst-Kalovski, 2010), respectively; results revealed that learners need from 4000- to 5000-word families to achieve 95% lexical coverage and 8000 to attain 98% coverage. However, it is not clear whether the coverage figures for reading may indeed be

transferred to viewing. Like listening, factors inherent to multimedia input, such as the presence of the imagery, may also affect viewing comprehension.

To date, only one study has empirically explored the relationship between lexical coverage and L2 viewing comprehension (Durbahn et al., 2020). Like Schmitt et al. (2011), the authors tested the words that occurred in the input. To that end, they sampled 146 (56%) words occurring in a 30-min documentary and tested them in an aural meaning-recall test. Learners' lexical coverage was calculated based on their results on the meaning-recall test. Additionally, the study also addressed the effect of imagery on viewing comprehension by assessing comprehension through questions designed to be answered by listening to the audio (audio-based), by paying attention to the imagery, and by listening to the audio (audio- and imagery-based) and viewing the imagery (imagery-based). One hundred and fourteen learners participated in the study. Results showed:

1. An almost medium-sized correlation between knowledge of the words in the documentary and viewing comprehension (r_s [94] = .39);
2. A positive, almost medium-sized correlation between audio-based questions and lexical coverage (r_s [94] = .36);
3. A small-sized correlation between imagery-based questions and lexical coverage (r_s [94] = .29), and
4. No significant correlation between lexical coverage and audio- plus imagery-based questions.

Findings indicated that lexical coverage for viewing documentaries is lower than the lexical coverage for unassisted reading but similar to informal listening (Schmitt et al., 2011). The nonsignificant correlation between audio- plus imagery-based questions and viewing comprehension may also suggest that vocabulary plays a different role in viewing comprehension than in listening comprehension because of the presence of imagery. Imagery may help learners access word meaning, guess the meaning of unfamiliar words from context, and complete the meaning of partially known words (Peters, 2019; Webb & Rodgers, 2009a). Additionally, viewing has been found to have lower lexical demands than reading (Webb & Rodgers, 2009a,b). Webb (2021), based on his discussion about lexical coverage and lexical profiling, indicates that 90% would be sufficient to understand audiovisual input. Finally, TV viewing may be easier to understand than listening because of the presence of imagery (Webb & Rodgers, 2009b). All the reasons provided above lead us to assume that transferring coverage figures from reading or listening to viewing comprehension may not be appropriate. One of the limitations of Durbahn et al.'s (2020) study is that they only examined the correlation between lexical coverage and viewing comprehension. Further research examining the effect that different levels of lexical coverage have on viewing comprehension is needed to better understand how the knowledge of words affects comprehension when the input is supported by imagery.

In summary, lexical coverage has been investigated by means of three methodological approaches:

1. By matching learners' scores on vocabulary (levels) tests with the profile of the text they are supposed to read or listen to;
2. Manipulating text by replacing words with nonwords to create different lexical coverage levels (Giordano, 2021; Hu & Nation, 2000) and
3. Corpus analyses (Webb & Rodgers, 2009a, 2009b).

The first approach can, in turn, be divided into: estimating the mastery of different frequency levels (Laufer & Ravenhorst-Kalovski, 2010; Noreillie et al., 2018) and assessing the knowledge of words that appear in target written or spoken texts (Schmitt et al., 2011; Durbahn et al., 2020). The use of VLTs and corpus analyses can be considered indirect ways of determining lexical coverage, whereas text manipulation can more accurately give an account of learners' lexical coverage. Therefore, more direct methods, such as text manipulation, are needed to investigate the effect of the degree of lexical coverage on viewing comprehension. Furthermore, it is worth noting that there is no consensus on what "adequate comprehension" is. What is adequate in one study may not be so in another (Nurmukhamedov & Webb, 2019) because adequate "may refer to different levels of comprehension in different contexts" (Laufer & Ravenhorst-Kalovski, 2010, p. 16). Consequently, results cannot always be compared.

Importantly, there may be factors involved in comprehension other than vocabulary knowledge, such as background knowledge (Pulido, 2007; Song & Reynolds, 2022; Schmitt et al., 2011) or inferring ability (Laufer, 2020), which make comprehension vary among learners when facing written and spoken input even if they have the same coverage. In several studies (Hu & Nation, 2000; Schmitt et al., 2011; van Zeeland & Schmitt, 2013, among others), students with even 100% coverage had poor comprehension scores. That is why Webb (2021) stresses that knowing 100% of the words in the input does not ensure 100% comprehension scores and that spoken and written texts are not always thoroughly understood at that level. Therefore, "while reaching target lexical coverage figures might indicate that text might be comprehensible, it does not ensure comprehension" (Webb, 2021, p. 281).

The Present Study

The present study aimed to investigate the effect of different levels of lexical coverage on TV viewing comprehension. The study adopted the same methodology as Hu and Nation (2000), van Zeeland and Schmitt (2013), Giordano (2020), and Song and Reynolds (2022). In other words, words occurring in the input are replaced with nonwords to create four levels of lexical coverage: 100%, 95%, 90%, and 80%. In line with the study by van Zeeland and Schmitt (2013), and to have a baseline, the present study involved the participation of L1 as well as L2 speakers of English. Most studies on the relationship between lexical coverage and viewing comprehension borrow figures from reading or listening. However, viewing presupposes advantages that single-channel modes of input, such as reading and listening, do not. Based on the Cognitive Theory of Multimedia Learning (Mayer, 2005), humans have a limited capacity to process information, but when that information is presented through two channels, auditory and visual channels, they can process it more efficiently. That is why treating viewing comprehension in the same way as reading and listening may not be appropriate. It is worth researching whether viewing is a construct in itself or a form of listening because of the role that imagery plays in comprehension. Additionally, as there is little research on the relationship between lexical coverage and viewing comprehension, there are no clear lexical goals set to understand viewing input as in reading research, for example. Findings from this study will contribute to knowledge of the relationship between lexical coverage and viewing comprehension in L1 and L2. Furthermore, they will help treat viewing as a construct different from reading and listening. They will also add to the understanding of how much vocabulary is needed to

comprehend audiovisual input, more specifically, documentaries. In that way, this study will help set learning goals in EFL programs and materials design.

The research question under investigation is: What is the effect of different lexical coverage levels on viewing comprehension?

Methodology

This study adopted a counterbalanced within-subjects experiment with four experimental conditions corresponding to four coverage levels: 100%, 95%, 90%, and 80%. This means that all participants took part in all experimental conditions but in a different order to control for order, topic effects, and individual differences. A within-participant design has the advantage that more power can be obtained without needing more participants (Godfroid, 2020).

Participants

Ninety-seven Spanish speakers whose L2 was English participated in the study. They were all undergraduate students from a university in Chile from the following programs: Bachelor in Language, English Language Pedagogy, English Literature and Linguistics, and Bachelor of Arts and Humanities. All participants were taking either English Language II or English Language IV courses, which means they had more than 5 hours a week of formal English language instruction plus other English language-related courses. They were selected on the basis of their scores on Webb et al.'s (2017) updated Vocabulary Levels Tests (uVLT) for two reasons:

- To identify whether they were familiar with at least the 2000 words and
- To match participants' vocabulary test scores with the text profile and ensure that the coverage levels were indeed 100%, 95%, 90%, and 80% and not lower.

Results from the uVLT indicated that 17 participants were not familiar with 1K or 2K; 2 participants had an L1 other than Spanish; and 2 participants' comprehension scores in all the video clips were too low (one or two correct answers), meaning that they either did not have the proficiency level adequate to understand the clips, which was not the case according to the uVLT, or that they were not taking the study seriously. Data from those learners were excluded, resulting in 76 L2 participants. The remaining participants were 61 females, 12 males, 2 agender, and 1 nonbinary. Their ages ranged from 18 to 40 years old ($M = 21.30$; $SD = 4.60$). Table 1 presents the descriptive statistics of the uVLT results after the exclusion of the participants mentioned above.

As in van Zeeland and Schmitt, data was also collected with native speakers of English. Sixty L1 learners took part in this study. However, only 40 L1 participants completed the experiment. They were invited via the Prolific platform, an online platform that finds and invites participants who match a specific study's requirements and pays them after they have completed all the tasks. Only L1 English speakers who were currently studying at university were invited to make the sample of L1 speakers as comparable as possible with L2 speakers. No specific regions were targeted. The L1 speaker participants were 25 women and 15 men whose ages ranged from 18 to 40 years old ($M = 22.53$; $SD = 5.14$). There were 39 undergraduate students and 1 graduate student. All participants were from the United Kingdom living in the United Kingdom, except for one who was living in Spain.

Table 1. Descriptive statistics of the updated vocabulary levels test (N = 76)

Levels	M	SD	Score		95% CI	
			Min	Max	Upper bound	Lower bound
1K	29.89	.31	29	30	29.97	29.82
2K	29.83	.38	29	30	29.92	29.74
3K	27.99	1.43	23	30	28.31	27.66
4K	27.99	2.72	18	30	28.61	27.37
5K	28.33	2.55	16	30	28.91	27.75

Table 2. Lexical frequency profile of the clips unaltered

Frequency bands	Deep water	Forest	Fresh water	Mountains
1K+ proper names and cognates, %	85.41	82.31	83.76	83.50
2K, %	13.52	16.97	15.5	15.82
>2K, %	1.07	0.72	0.74	0.68

Video Excerpts and Viewing Comprehension Test

The audiovisual texts used were four clips of approximately 2 min each (Table 2) of the first season of the documentary *Planet Earth* (2006), episodes: Deep Water, Forests, Fresh Water, and Mountains. The same documentary as Rodgers (2018) and Peters (2019) was selected for comparability purposes. Documentaries were chosen because they have been found to contain more word-imagery matching than narrative TV genres (Rodgers, 2018). Additionally, in the background questionnaire, 100% of the L2 participants reported not having seen the documentary before.

As this study's purpose was to explore the extent to which lexical coverage affects viewing comprehension, it was important to determine if the clips were similar in terms of comprehensibility. We asked six highly proficient L2 and eight L1 speakers to rate the clips. They watched the unaltered clips and ranked them according to their level of comprehensibility. Eight raters indicated that the videos were equally comprehensible. Five of the 13 participants indicated that the clips had different levels of comprehensibility. However, there was no agreement on the hierarchy of comprehensibility, which led us to assume that none of the clips was particularly more challenging than the others.

As the clips needed reworking by recording the voice over to create the different levels of lexical coverage, they were transcribed with the help of the online tool Happy Scribe (www.happyscribe.com). They were cleaned by transforming all the contractions to their full forms (Rodgers, 2013). Next, they were submitted to the free software, AntWordProfiler (Anthony, 2014; Table 2) to compute the lexical profile of the clips. Because the clips were short, some sentences were added so that they had a similar number of tokens. Sentences were also added to clarify or complement some of the information. For example, in the Fresh Water video, it said: *The skin of some species contains a powerful poison; these salamanders tend to be slow-moving and have bright warning coloration to advertise their harmfulness. Free from competition, these giants can dine alone.* The sentence about their warning coloration was added to complement the information about the poison in their skins. Following Hu and Nation (2000), the texts were submitted to a process of simplification to make sure all the words were within the 2K most frequent word families. Words that fell over the 2K level were replaced with words of a higher frequency. For example, *pine needles* was replaced with

pine's sharp pins. A native speaker checked the clips for the naturalness of the language. Results from the lexical frequency profile after the modifications showed that approximately 99% of the tokens in the videos belonged to the 2000 most frequent word families (provided proper names, such as *Japan* or *Indonesia*, and English to Spanish cognates, such as *opponent* or *coral*, are known). A list of potential cognates was created. A native English speaker, the same person who did the voice-over, was asked to record the words. Subsequently, a group of five native Spanish speakers was informally asked to recognize the words from the list. A word was considered a cognate when it was recognized by all five Spanish speakers. The rating procedure plus the lexical frequency profile showed that the four clips were very similar in terms of topic difficulty and vocabulary load, respectively.

We then created 100%, 95%, 90%, and 80% coverage levels by replacing 0%, 5%, 10%, and 20% of the words in the clips with nonwords, respectively (Table 3). The replaced words were generally words from the 3K to 22K or off-the-list words and were the words that could not be replaced in the process of simplification. As there were not enough items of this type, words were taken randomly, using the free online software www.random.org, from 2K to 1K (in the case of 80%, e.g.). As in Hu and Nation (2000), if a word occurred more than once in the text, it was replaced with the same no-word on every occasion. The nonwords were taken from the ARC Nonword Database (Rastle, Harrington, & Coltheart, 2002), which is the same database that was used by van Zeeland and Schmitt (2013). All nonwords replaced into the passages complied with the phonological and orthographic rules of English, which is the purpose of the database, and were five to eight phonemes long. They sounded or looked like the part of speech of the words they replaced, and participants could easily identify their part of speech, for example: *contain → shaint* (Appendix A). As the words replaced were mostly low-frequency words and the ones that were high-frequency words were selected at random, each version of the clip would realistically represent the problems that L2 learners may encounter when viewing videos.

Once the items were replaced, the narration for the videos was read and recorded from the scripts by a voice talent at a rate of 118.85 words per minute. To control for individual differences such as pronunciation, accent, and speaking style, the four clips were narrated by the same speaker in standard English (a Canadian L1 English male speaker). The original sound editing was carried out using the software Audacity, and the video editing was done using the software Cyberlink PowerDirector 18.

The present study employed a counterbalanced within-subjects design, where four levels of lexical coverage, 100%, 95%, 90%, and 80%, were created for each topic (Table 4). Each clip was recorded four times for each level of lexical coverage. Later, the videos were combined in such a way that four playlists were created containing one video of each topic with one different level of lexical coverage, all of them in a different

Table 3. Number of words and nonwords in the four clips

Clips	Lexical coverage (%) Number of tokens Number of nonwords in text								Clip length
	100	0	95	5	90	10	80	20	m: s
Deep Water	281	0	266.5	14.05	252.9	28.1	224.8	56.2	2:23
Forests	277	0	263.15	13.85	249.3	27.7	221.6	55.4	2:07
Fresh Water	271	0	257.45	13.55	243.9	27.1	216.8	54.2	2:14
Mountains	297	0	282.15	14.85	267.3	29.7	237.6	59.4	2:26

order, as shown in Table 4. The playlists were saved as an mp4 file and uploaded on the testing platform Playposit, where the questions were added. Participants were then assigned to one of the four playlists.

The questions were created as follows: A large set of questions was created based on the information in the clips, which resulted in 80 multiple-choice questions. Next, the questions were classified according to the information needed to answer them: verbal information and/or visual information. This resulted in 36 audio-based and 44 audio-plus imagery-based questions. Because of the length of the clips, all the questions were literal wh-questions relating to the central topics of the clips organized in chronological order. They contained three distractors that were related to the topics of the clips, a correct answer, and an "I don't know" option to avoid guessing. The questions were written in the participants' first language (English for L1 speakers and Spanish for L2 speakers). They were added after each clip to the testing platform Playposit. After the questions were created, two researchers rated the questions as audio- or audio- plus imagery-based questions. The questions were piloted with a group of 12 volunteers that included researchers, teachers, and proficient L2 speakers. Results revealed that some questions needed rewording, and questions with low scores were excluded. The final set of questions included seven audio-based and seven audio- plus imagery-based questions per clip (N = 56) constituted the comprehension test. A second pilot study with a representative sample of 10 EFL university students was conducted to ensure that the questions were clear, had no obvious answers, had only one correct answer, and that the format on the platform was appropriate. Results showed that pilot participants had, on average, 82.78% (SD = 9.77) correct answers.

We examined Cronbach's α estimate of the internal consistency of each playlist. Reliability coefficients were calculated with the actual scores from L1 and L2 participants. A Cronbach's α of .70 is often put forward for acceptable reliability (Cronbach, 1951). Given the low number of questions, the playlists had an acceptable level of reliability at 100%, albeit with a lower coefficient for playlist D, as can be seen in Table 5. Taber (2017) explains that Cronbach's α should be interpreted according to the context of a particular study considering the total number of items. In the case of this study, each clip had 14 questions, which may be the reason for the medium reliability coefficient. When the reliability coefficient was computed for each playlist, the α

Table 4. Lexical coverage distribution per video topic

Playlists	Videos and lexical coverages, %			
A	Deep Water 100	Forest 95	Fresh Water 90	Mountain 80
B	Forests 80	Deep Water 90	Mountains 95	Fresh Water 100
C	Fresh Water 95	Mountain 90	Forests 100	Deep Water 80
D	Forests 90	Fresh Water 80	Deep Water 95	Mountains 100

Table 5. Reliability coefficients (Cronbach's α) at 100% coverage and total per playlist

Playlists	100%	Total
A	.809	.860
B	.618	.844
C	.693	.813
D	.417	.697

Note: Clips or playlists were considered reliable when α was >.70

increased and each playlist was reliable as a whole. It is worth mentioning that reliability coefficients for listening are generally low compared to other skills (Plonsky & Derrick, 2016). Despite the lower reliability coefficient in Playlist D, a simple one-way analysis of variance revealed that the comprehension scores from the four different playlists were not significantly different, $p > .05$. To find the full report of reliability coefficients per clip and coverage levels, see Appendix C.

Updated Vocabulary Levels Test

To ensure that the L2 participants were familiar with the 2000 most frequent word families and would be able to understand the words occurring in the videos, they took an online version of Webb et al.'s (2017) updated VLTs created on Google Forms (Figure 1). The test contains 50 discrete, selective, context-independent clustered questions aiming at assessing the form-meaning link. Each cluster contains six words and three definitions that have to be matched with the words. In all, 150 words were assessed, 30 per level, from the first five frequency bands sampled from Nation's (2012) British National Corpus/Corpus of Contemporary American English word lists. The test designers suggest a cutoff score of 29 of 30 correct answers in the first three and 24 in the last two levels. As seen in Table 1, all learners were familiar with the 1K and 2K levels, which suggests they could understand the input from the clips. The test had a medium reliability coefficient ($\alpha = .680$).

Procedure

Due to the COVID-19 pandemic, the entire study was designed to be conducted via the Zoom platform. L2 students from eight intact classes were informed about the purpose of the study and asked to participate. They were also informed about the instructions for the activities. They received the links to the background questionnaire that also contained the informed consent and the online updated VLT. Approximately 1 month later, they received the link to the intervention, which consisted of the video clips and comprehension test. They were warned that some clips contained nonwords to prevent them from taking extra time trying to guess the meaning of the words. Each clip was played twice. After the first clip was played for the first time, the video was paused automatically, and all questions were displayed on the left part of the screen (Figure 2). At this point, participants could only read the questions but not answer them so that they could have an idea of what was going to be asked. The video was paused until the

1

	choice	computer	garden	photograph	price	week
cost	○	○	○	○	○	○
picture	○	○	○	○	○	○
place where …	○	○	○	○	○	○

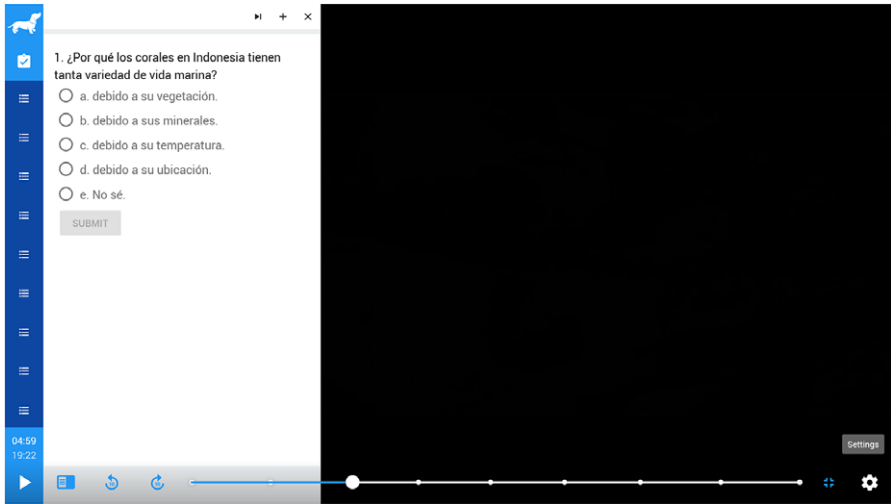**Figure 1.** Example of an item of the updated vocabulary levels test.

Figure 2. Example of an item of the comprehension test.

participants read all the questions and clicked "Continue." Once the video was played for the second time, it was paused again, and the questions popped out one by one. Participants could not rewind or skip questions. When all the questions were answered, the video resumed. This process was repeated with each clip until they completed the comprehension test. Each playlist with the four clips plus the pauses lasted approximately 20 min. The participants could not fast-forward the video, and the video automatically stopped if they changed the tab or minimized the window. Participants were connected to Zoom throughout both sessions to ask questions, and their progress was tracked through the Playposit platform. They took, on average, 64 min as indicated on the platform.

The procedure for the L1 participants was identical to that for the L2 participants, except that they did not take the updated VLT. Their consent form was embedded in the participants' demographic information. On average, L1 speakers took 43 min and 42 s.

Scoring

All the tests were scored dichotomously. If an answer was correct, it was awarded 1 point and 0 if not. Both platforms, Google Forms, and Playposit, allowed for inputting the answer keys before participants took the test, and consequently, the test results could be downloaded in CSV format.

Analysis and Results

We calculated the comprehension scores and grouped the participants according to these scores (from 0 to 14 of 14) at the different levels of lexical coverage (viz. 100%, 95%, 90%, and 80%). In other words, we calculated how many participants had 14, 13 points, etc., at each coverage level. Figures 3 and 5 show that there is a visible downward trend regarding the comprehension scores at the different levels of lexical coverage for both L1 and L2 speaker participants (Tables 6 and 7).
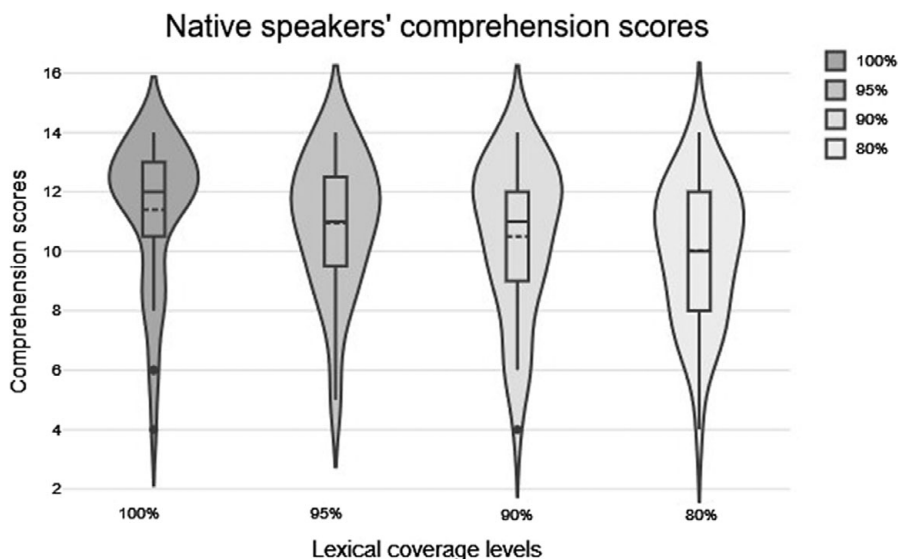
Figure 3. Native speakers' distribution of results per lexical coverage level.

Table 6. Number of L1 participants in each score point of the comprehension test (N = 40)

Comprehension scores	Lexical coverage			
	100%	95%	90%	80%
14	6	5	3	2
13	9	5	6	2
12	13	9	10	9
11	2	7	3	6
10	1	4	6	6
9	3	5	4	4
8	3	2	2	4
7	0	1	1	5
6	2	0	4	1
5	0	2	0	0
4	1	0	1	1
3	0	0	0	0
2	0	0	0	0
1	0	0	0	0
Mean score	11.40	10.95	10.50	10.03
SD	2.46	2.29	2.56	2.31
Lower bound	10.61	10.21	9.68	9.28
Upper bound	12.19	11.68	11.32	10.77

L1 speakers

Results from native speaker participants indicated that, even with 100% coverage, only six L1 participants obtained the maximum score of 14. The most recurring score (mode) was 12 points on the comprehension test, with 13 of 40 participants obtaining that score. Table 6 shows the number of participants in each comprehension score at

Table 7. Number of L2 participants in each score point of the comprehension test (N = 76)

Comprehension scores	Lexical coverage			
	100%	95%	90%	80%
14	17	7	3	2
13	20	17	14	6
12	14	18	13	13
11	8	13	16	21
10	7	5	13	13
9	1	9	9	10
8	3	5	2	3
7	4	2	2	3
6	0	0	3	1
5	2	0	0	3
4	0	0	1	0
3	0	0	0	1
2	0	0	0	0
1	0	0	0	0
Mean score	11.77	11.36	10.78	10.29
SD	2.24	1.85	2.05	2.13
Lower bound	11.27	10.93	10.31	9.80
Upper bound	12.29	11.77	11.24	10.77

each lexical coverage percentage, whereas Figure 3 shows the distribution of participants' scores per lexical coverage level.

To answer whether lexical coverage affected the comprehension of L1 speakers, we ran the non-parametric Friedman test because the data was not normally distributed. The results showed that the degree of lexical coverage had a significant effect on comprehension, $\chi^2(3) = 13.077$, $p = .004$, $w = .019$, but with a small effect size. Pairwise comparisons were performed with a Bonferroni correction for multiple comparisons. Post hoc analysis revealed statistically significant differences in comprehension levels between 100% and 90% coverage ($p = .044$) and between 100% and 80% coverage ($p = .019$). All other comparisons were not statistically significant (Figure 4).

L2 speakers

Results from non-native speaker participants indicated that, in the 100% coverage level, the most recurring score (mode) was 13 of 14 points on the comprehension test, with 20 of 76 participants obtaining that score. Most participants obtained between 12 and 14 correct answers. This indicates that with 100% lexical coverage, most participants were able to obtain between 85.7% and 100% correct answers on the comprehension test. In the 95% coverage, the highest scores were between 11 and 13. Comprehension trended downward in the other two lower levels. Table 7 shows the number of participants in each comprehension score at each lexical coverage percentage, whereas Figure 5 shows the distribution of participants' scores per lexical coverage level.

We also ran a Friedman test for the L2 data. The results showed that the degree of lexical coverage had a significant effect on comprehension, $\chi^2(3) = 30.759$, $p < .001$, $w = .135$, but with a small effect size. Pairwise comparisons were performed with a Bonferroni correction for multiple comparisons. Post hoc analysis revealed statistically significant differences in comprehension levels between 100% and 90% lexical coverage
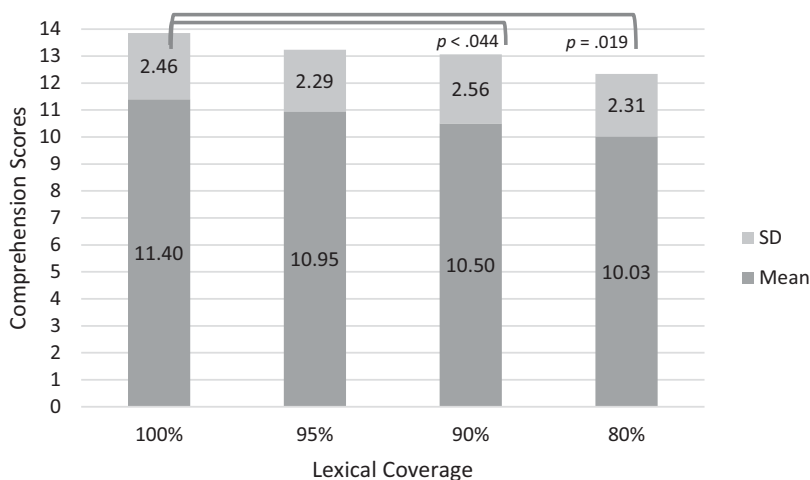
Figure 4. Mean comprehension scores by L1 speakers on the four coverage levels with a Bonferroni correction.
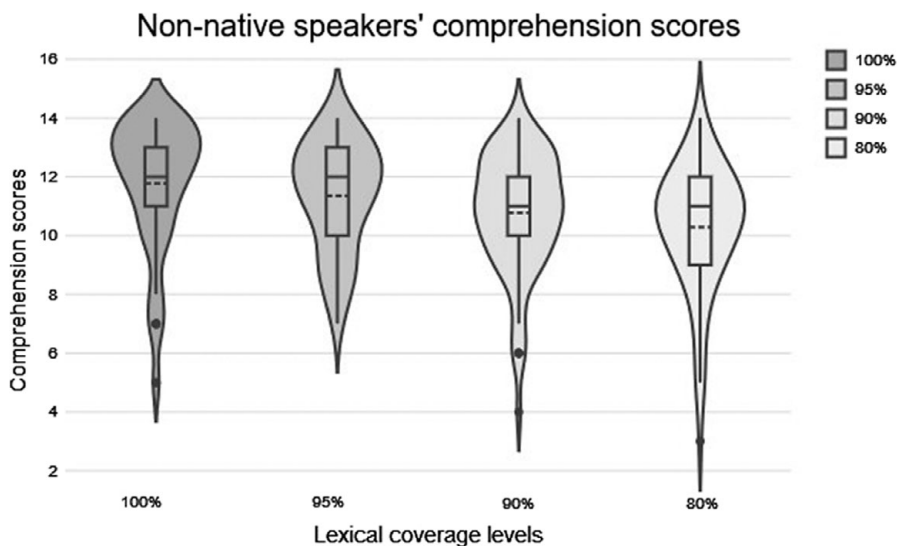

Figure 5. Non-native speakers' distribution of results per lexical coverage level.

($p = .007$), between 100% and 80% coverage ($p < .001$), and between 95% and 80% coverage ($p = .015$). All other comparisons were not significant (Figure 6).

Because non-native speakers' comprehension scores were surprisingly higher than native speakers, an additional statistical analysis was conducted to examine whether the difference was statistically significant. Comprehension scores were not normally distributed as assessed Shapiro-Wilk's test ($p < .05$). Therefore, a Mann-Whitney U test was run. Distributions of the comprehension scores for native and non-native speakers were similar, as assessed by visual inspection. Results indicated that
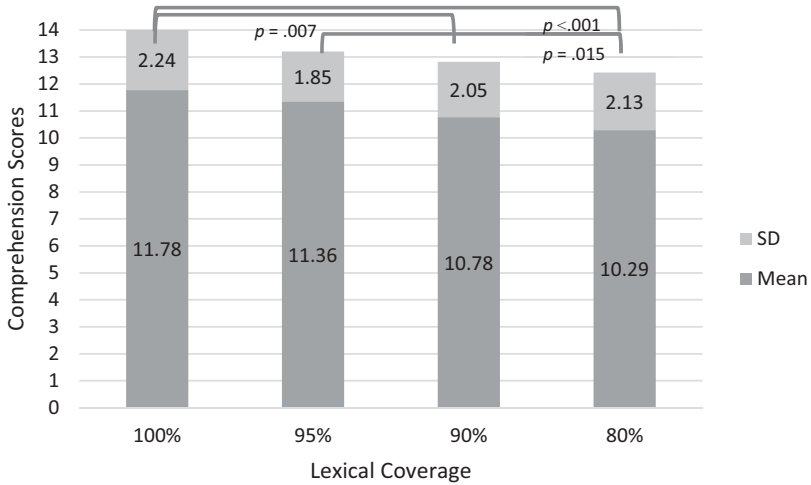
Figure 6. Mean comprehension scores by L2 speakers on the four coverage levels with a Bonferroni correction.

comprehension scores were not statistically significantly different between L1 and L2 speakers at the different levels of lexical coverage, 100%, $U = 1372.5$, $z = -.873$, $p = .383$; 95%, $U = 1386$, $z = -.789$, $p = .430$; 90%, $U = 1479$, $z = -.241$, $p = .810$; 80%, $U = 1417.5$, $z = -.603$, $p = .546$, using an exact sampling distribution for U (Dineen & Blakesley, 1973).

An additional analysis was carried out to compare the scores of the two different types of questions. Because the analysis is beyond the scope of this study, we did not include the results in this section. However, they can be found in Appendix B.

Considering the results from this study and following Laufer and Ravenhorst-Kalovski (2010), the cutoff score for adequate comprehension was conceptualized as the point at which more learners are likely to comprehend well. In this study, the most recurrent score (mode) in the native speakers' group at 100% lexical coverage (baseline) was 12 of 14 points. At this cutoff score, more than half of the native-speaker participants (25 of 40) had adequate comprehension. Regarding L2 participants, most learners received 13 at 100% coverage, but 12 is the point at which more than 50% of the participants reached adequate comprehension, altogether 51 of 76 participants. Additionally, it was observed that at 80% lexical coverage, the mean comprehension score was 10 of 14. At this point, a good proportion of participants could understand the clips. With a more lenient criterion, the threshold of adequate comprehension could also be set at 10 points, which is the lowest mean comprehension score of all the coverage percentages. In conclusion, we suggest two cutoff scores for viewing comprehension: one optimal reflected in the score of 12 and one minimal reflected in the score of 10 of 14 points.

Discussion

This study aimed to describe the extent to which the degree of lexical coverage affects L2 viewing comprehension. Four excerpts from the documentary *Planet Earth* were manipulated to create four different levels of lexical coverage, 100%, 95%, 90% to 80% coverage. The results suggest that the degree of lexical coverage plays a significant

role in viewing comprehension of short documentaries, as a steady decrease in comprehension scores was observed when the rate of unknown words increased. This was the case for both L1 and L2 speakers. Our findings lend support to research on the relationship between lexical coverage and reading (Hu & Nation, 2000; Laufer & Ravenhorst-Kalovski, 2010; Schmitt et al., 2011) listening (Stæhr, 2009; van Zeeland & Schmitt, 2013), and viewing (Durbahn et al., 2020), which have also found that the more known words in a text, the better the rate of comprehension, regardless of the methodology employed in the studies.

There were high means in all coverage levels of at least 10 of 14 points (Figure 7). Comprehension of the audiovisual input was significantly better at 100% coverage than at lower coverage levels (i.e., 90% and 80% in both groups), with means over 11 points in L1 and L2. Interestingly, not all the participants had the maximum score at 100% level. This is reminiscent of previous research (Hu & Nation, 2000; Webb, 2021) that has found that knowing all the words in the input does not ensure 100% correct answers on a comprehension test, neither in L1 nor L2. However, 100% lexical coverage still led to high levels of comprehension. These findings support previous research comparing the role that vocabulary plays in reading (Hu & Nation, 2000) and listening (Giordano, 2020; van Zeeland & Schmitt, 2013), as discussed in the literature review.

The high means on the comprehension test at the four levels of coverage can be explained by the following reasons. First, Mayer's (2009) theory of multimedia learning postulates that learning is fostered when learners process information simultaneously through visual and auditory channels. It is worth noting that, although Mayer's theory is about learning and not about comprehension, it suggests that when words and imagery are integrated effectively, they can work synergistically to help learners cope with the information that was missing due to the nonwords and comprehension was
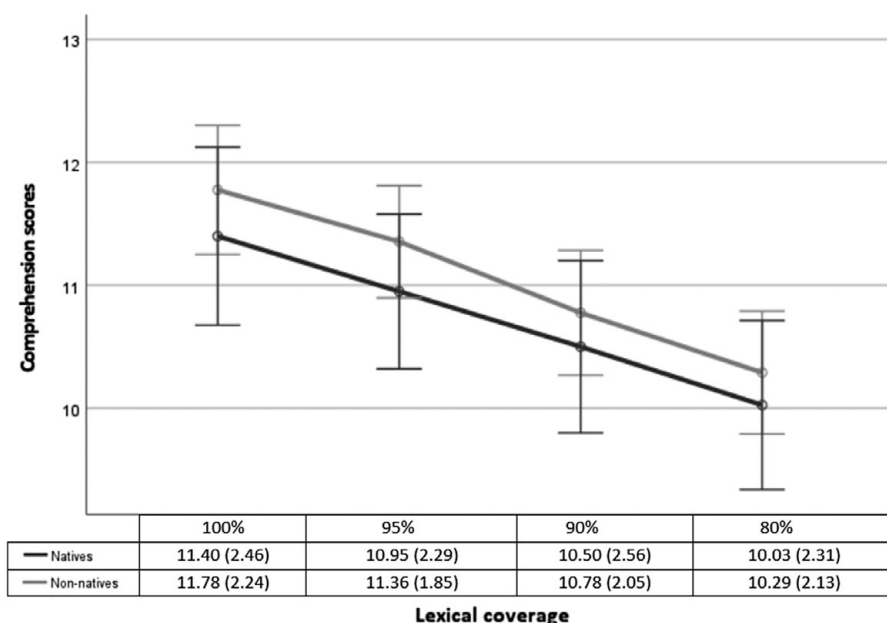

	100%	95%	90%	80%
──── Natives	11.40 (2.46)	10.95 (2.29)	10.50 (2.56)	10.03 (2.31)
──── Non-natives	11.78 (2.24)	11.36 (1.85)	10.78 (2.05)	10.29 (2.13)

Lexical coverage

Figure 7. Mean comprehension scores by both L1 and L2 speakers. Error bars show the mean ± 1 SD.

improved as a result. Second, research has found that the way in which imagery cooccurs with the spoken presentation of vocabulary, as in television programs, may be conducive to learning (Rodgers, 2018). Finally, this aural-visual cooccurrence is found more often in documentary television.

Caution must be taken when interpreting the scores since the high means also carried high degrees of variation in comprehension scores across all levels of lexical coverage. This suggests that some learners were able to cope with unknown vocabulary, even at low levels of lexical coverage, as in the case of the 80% coverage point. Likewise, our findings suggest that some learners had difficulties understanding the input even when they had a high level of lexical coverage, as in the case of the 95% and 100% coverage levels. Durbahn et al. (2020) also found considerable variation in the viewing comprehension scores at each coverage point, except at the 99% level, which was probably due to the reduced number of learners. It seems that audiovisual input tends to create variation in the comprehension scores and that the role vocabulary plays in viewing comprehension might be affected by other factors, such as imagery. This may be explained by the almost medium-size correlation between lexical coverage and viewing comprehension found in Durbahn et al. (2020). Schmitt et al. (2011) stated that with so much variation, it would be difficult to predict comprehension based merely on vocabulary knowledge. The contextual cues that imagery provides may help learners expand their lexical repertoire and complement the knowledge of partially known words (Durbahn, et al., 2020; Peters, 2019; Peters & Webb, 2018; Rodgers, 2018; Webb & Rodgers, 2009a, 2009b), which would result in an advantage over reading and listening. In the same vein as previous research (Durbahn et al., 2020), results from this study seem to suggest that when imagery is present, learners tend to rely less on the spoken input and more on the imagery or at least confirm what is being said by paying attention to the imagery (Durbahn et al., 2020).

The coverage necessary to view documentaries will depend on what is considered adequate comprehension. Regardless of the text type, different degrees of comprehension may be required for different purposes, for example, to read for pleasure, to listen to informal texts, etc.. Hence, previous research has defined "adequate" comprehension differently. In this study, we suggest two thresholds for adequate comprehension: *optimal* set at a score of 12 and *minimal* set at a score of 10 of 14 points. With an optimal adequate comprehension cutoff score of 12, the coverage level at which there was a larger number of learners who reached optimal adequate viewing comprehension is 95%. As shown in Tables 6 and 7, at a 95% coverage level, a cutoff score of 12 of 14 results in 52.6% of participants (47.5% of L1 and 55.3% of L2 speakers) having optimal adequate comprehension. With a minimal adequate comprehension cutoff score of 10 of 14, 80% is the coverage level where more than half of the participants obtained a score of 10 or more. At this coverage level, 25 L1 speakers (62.5%) and 55 L2 speakers (72.4%) meet the threshold for minimal comprehension. Additionally, 95% was statistically significantly different than 80% (L2 speakers) coverage level. In conclusion, we suggest an optimal lexical coverage of 95% and a minimum lexical coverage of 80% to achieve a viewing comprehension score of 85% and 70%, respectively, in audio-based and audio- plus imagery-based questions from short documentaries.

Although not significantly different, L2 speakers' scores were higher than L1 speakers' scores, and the downward trend is clearer as well. Those findings contradict results from van Zeeland and Schmitt (2013), where native speakers of English had overall higher results than non-native speakers. There are several possible reasons for this difference. One reason might be that L2 speakers were used to being exposed to input followed by answering comprehension questions. Given that it is common in

language classroom settings (a context all the L2 speakers were currently a part of), it is a situation that they expect and are ready to carry out, and thus may have paid more attention to the input. Second, it might be that L2 speakers are better trained to cope with words they do not know (van Zeeland and Schmitt, 2013), and, in that way, they might have used compensatory strategies to compensate for the gaps in comprehension. Another factor is that all L2 participants were undergraduate students of a program related to English-language teaching or learning, and despite their different levels of proficiency, they were all trained in listening and test-taking strategies, whereas the L1 participants pool was much more varied and did not necessarily involve students of language-related programs. Another plausible reason could be related to the data collection procedure. Although non-native speakers' data collection procedure was carried out in a real-time Zoom session, in which learners did all the tasks while still connected, native speakers' data collection was carried out through a data collection platform, in which little control could be exerted other than the test-taking time. Finally, the difference could be related to sample size. L2 speakers doubled the L1 speakers, and the larger the sample size, the larger the effects show (Field, 2013).

Pedagogical implications

This study had several pedagogical implications. First, findings showed that for learners to comprehend short documentaries, there needs to be a clear emphasis on improving their lexical knowledge in the L2 until they reach at least 80% lexical coverage. To do so, learners would need to be familiar with the most frequent 1000-word families, as it is the band that covers >80% of movies and television programs, according to Webb and Rodgers (2009a, 2009b). Second, these findings may be important for teachers and course designers in setting vocabulary goals. Results may also be useful for materials designers to create audiovisual input that goes according to the learners' proficiency level. Thus, if graded reading exists, so can graded viewing. Next, the Cognitive Theory of Multimedia Learning (Mayer, 2009) suggests that instructional material designed, taking into consideration how the human mind works, has better chances of leading to meaningful learning than material not designed with that purpose. Therefore, we stress the need to treat viewing as a construct different from listening, considering the role imagery plays. Last, as lexical coverage may be lower for viewing documentaries than for reading or listening, it seems sensible to suggest using short clips at lower levels of proficiency to increase learners' confidence in the L2 language.

Limitations

There were a few limitations to the current study. The first deals with the control of the experiment. Because the study was conducted online because of pandemic reasons, we had limited control over the data collection. The second limitation deals with the input. The audiovisual input used in this study consisted of four short clips from the documentary series *Planet Earth.* Imagery in documentaries has been found to have a higher degree of word-imagery co-occurrence than other narrative television programs (Rodgers, 2018). Consequently, findings from this study may only be generalizable to that specific type of audiovisual input and may not necessarily reflect the lexical coverage to understand other television genres. More research with other TV genres should be conducted if we want to generalize the present findings. The third

limitation was related to ecological validity. It would be difficult for teachers to replicate this study since reworking audiovisual input, even if they are as short as 2 min, is time-consuming and can be expensive if the voice-over is done by an experienced voice talent. The last limitation deals with the low-reliability coefficient found in Playlist D. Low reliability suggests that the clips within Playlist D do not consistently measure the same underlying construct. This reduces the internal consistency of the measurement instrument. Results from this Playlist might not be as reliable or applicable to the broader population or context.

Conclusion

This study investigated the effect of different levels of lexical coverage on viewing comprehension. Results indicated that the degree of lexical coverage affects comprehension of documentaries and that as the number of unknown words increases, the comprehension scores decrease. Results also showed that the scores at 100% lexical coverage level were significantly higher than at 90% and 80% levels in the L1 and L2 groups and that L2 learners' scores at 95% lexical coverage were significantly higher than the scores at 80%.

   The findings also revealed that the lexical coverage necessary for viewing a short documentary might be lower than for reading and listening. This is because, with coverage as low as 80%, the mean comprehension scores were >70% in both groups. We suggest an optimal lexical coverage of 95% to achieve an optimal adequate viewing comprehension score of at least 85.7%, and a minimal lexical coverage of 80% to achieve a minimal adequate viewing comprehension score of 71.4% or more. Finally, results clearly show how imagery aids comprehension causing viewing to differ from reading and listening.

References

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

Carver, R. P. (1994). Percentage of unknown vocabulary words in text as a function of the relative difficulty of the text: Implications for instruction. *Journal of Reading Behavior 26(*4), 413–437.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297–334. doi:10.1007/bf02310555

Durbahn, M., Rodgers, M., & Peters, E. (2020). The relationship between vocabulary and viewing comprehension. *System*, *88*, 102166. https://doi.org/10.1016/j.system.2019.102166

Dineen, L. C., & Blakesley, B. C. (1973). Algorithm AS 62: Generator for the sampling distribution of the Mann-Whitney U statistic. *Applied Statistics*, *22*, 269–273.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics and sex and drugs and rock'n'Roll*. Sage.

Giordano, M. J. (2021). *Lexical coverage in dialogue listening*. Language Teaching Research. https://doi.org/10.1177/1362168821989869

Godfroid, A. (2020). *Eye Tracking in Second Language Acquisition and Bilingualism.* https://doi.org/10.4324/9781315775616

Hu, M., & Nation, I. S. P. (2000). Vocabulary density and reading comprehension. *Reading in a Foreign Language*, *13*(1), 403–430.

Jeon, E.H., & Yamashita, J. (2014). L2 reading comprehension and its correlates: A meta-analysis. *Language Learning*, *64*, 160–212.

Laufer, B. (1989). What percentage of text lexis is essential for comprehension? In C. Lauren, and M. Nordman (eds.): *Special language: From humans thinking to thinking machines*. Clevedon: Multilingual Matters, pp. 316–323.

Laufer, B. (1992). Reading in a foreign language: How does L2 lexical knowledge interact with the reader's general academic ability? *Journal of Research in Reading*, 15(2), 95–103. https://doi.org/10.1111/j.1467-9817.1992.tb00025.x

Laufer, B. (2020). Lexical coverages, inferencing unknown words and reading comprehension: How are they related? *TESOL Quarterly*, 54(4), 1076–1085. https://doi.org/10.1002/tesq.3004

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.

Mayer, R. E. (2005). *The Cambridge Handbook of Multimedia Learning*. University of Cambridge.

Mayer, R. E. (2009). *Multimedia learning* 2nd ed. Cambridge: Cambridge University Press.

Mayer, R. E., Lee, H., & Peebles, A. (2014). Multimedia learning in a second language: A cognitive load perspective. *Applied Cognitive Psychology*, 28, 653–660.

Nation, I. S. P. (1983). Testing and teaching vocabulary. *Guidelines*, 5, 12–25.

Nation, I. S. P. (2001). *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139524759

Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *The Canadian Modern Language Review*, 63, 59–82.

Nation, I. S. P. (2012). Information on the BNC/COCA lists. Retrieved from http://www.victoria.ac.nz/lals/about/staff/publications/paul-nation/Informationon-the-BNC_COCA-word-family-lists.pdf

Noreillie, A. S., Kestemont, B., Heylen, K., Desmet, P., & Peters, E. (2018). Vocabulary knowledge and listening comprehension at an intermediate level in English and French as foreign languages: An approximate replication study of Stæhr (2009). *ITL–International Journal of Applied Linguistics*, 169(1), 212–231.

Nurmukhamedov, U. & Webb. S. (2019). Lexical coverage and profiling. *Language Teaching*, 52(02), 188–200. https://doi.org/10.1017/s0261444819000028

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

Perfetti, C. & Stafura. J. (2014). Word Knowledge in a Theory of Reading Comprehension. *Scientific Studies of Reading*, 18(1), 22–37. https://doi.org/10.1080/10888438.2013.827687

Peters, E. (2019). The effect of imagery and on-screen text on Foreign Language Vocabulary Learning from audiovisual input. *TESOL Quarterly*, 53(4), 1008–1032. https://doi.org/10.1002/tesq.531

Peters, E. (2019). Factors affecting the learning of single-word items 1. *The Routledge Handbook of Vocabulary Studies*, 125–142. https://doi.org/10.4324/9780429291586-9

Peters, E., & Webb, S. (2018). Incidental vocabulary acquisition through viewing l2 television and factors that affect learning. *Studies in Second Language Acquisition*, 40, 1–27

Peters, E., Velghe, T., & Van Rompaey, T. (2015, May). *A post-entry English and French vocabulary size for Flemish learners*. Paper presented at EALTA. Copenhagen, Denmark.

Plonsky, L., & Derrick, D.J. (2016). A meta-analysis of reliability coefficients in second language research. *Modern Language Journal*, 100(2), 538–553. https://doi.org/10.1111/modl.12335

Pulido, D. (2007). The Relationship Between Text Comprehension and Second Language Incidental Vocabulary Acquisition: A Matter of Topic Familiarity? 57(Suppl 1), 155–199. https://doi.org/10.1111/j.1467-9922.2007.00415.x

Qian, D. D. (1999). Assessing the roles of depth and breadth of vocabulary knowledge in reading comprehension. *The Canadian Modern Language Review*, 56(2), 282–307.

Rastle, K., Harrington, J., & Coltheart, M. (2002). 358,534 nonwords: The ARC Nonword Database. *Quarterly Journal of Experimental Psychology* 55A: 1339–1362.

Rodgers, M. P. H. (2013). *English language learning through viewing television: An investigation of comprehension, incidental vocabulary acquisition, lexical coverage, attitudes, and captions*. New Zealand: Victoria University of Wellington. Unpublished doctoral thesis.

Rodgers, M. P. H. (2018). The images in television programs and the potential for learning unknown words: The relationship between on-screen imagery and vocabulary. *ITL - International Journal of Applied Linguistics*, 169, 192–213.

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., & Clapham, C. (2001). Developing and exploring the behaviour of two versions of the vocabulary levels test. *Language Testing*, *18*(1), 55–88. https://doi.org/10.1177/026553220101800103

Song, T., & Reynolds, B. L. (2022). The effects of lexical coverage and topic familiarity on the comprehension of L2 expository texts. *TESOL Quarterly*, *56*(2), 763–774. https://doi.org/10.1002/tesq.3100

Stæhr, L. S. (2008). Vocabulary size and the skills of reading, listening and writing. *Language Learning Journal*, *36*, 139–152.

Stæhr, L. S. (2009). Vocabulary knowledge and advanced listening comprehension in English as a foreign language. *Studies in Second Language Acquisition*, *31*(4), 577–607.

Taber, K. S. (2017). The use of Cronbach's alpha when developing and Reporting Research Instruments in science education. *Research in Science Education*, *48*(6), 1273–1296. https://doi.org/10.1007/s11165-016-9602-2

van Zeeland, H. & Schmitt, N. (2013). Lexical coverage in L1 and L2 listening comprehension: The same or different from reading comprehension? *Applied Linguistics*, *34*(4), 457–479.

Webb, S. (2021). Research investigating lexical coverage and lexical profiling: What we know, what we don't know, and what needs to be examined. *Reading in a Foreign Language*, *33*(2), 287–302.

Webb, S., & Rodgers, M. P. H. (2009a). The lexical coverage of movies. *Applied Linguistics*, *30*(4), 407–427.

Webb, S., & Rodgers, M. P. H. (2009b). Vocabulary demands of television programs. *Language Learning*, *59*(2), 335–366.

Webb, S., Sasao, Y., & Balance, O. (2017). The updated Vocabulary Levels Test: Developing and validating two new forms of the VLT. *ITL - International Journal of Applied Linguistics*, *168*(1), 34–70.

Zhang, S., & Zhang, X. (2020). The relationship between vocabulary knowledge and L2 reading/listening comprehension: A meta-analysis. *Language Teaching Research*, *26*(4), 696–725. https://doi.org/10.1177/1362168820913998